矩阵分析

冯良贵　胡庆军　编著

国防科技大学出版社

矩阵分析

矩阵分析

封面设计：燕军

策划人：耿筠

ISBN 978-7-81099-767-6



定价：29.00元

矩 阵 分 析

冯良贵 胡庆军 编著

国防科技大学出版社
长沙

内容简介

本书较系统地介绍了矩阵分析的基本内容、方法及其在相关学科中的应用。全书共分七章，主要内容包括：线性变换的矩阵表示、矩阵分解、矩阵的广义逆、矩阵范数与矩阵级数、矩阵函数及其计算、函数矩阵的微积分及应用、特征值估计等。内容力求阐述简明、推导严谨，并配有适量的习题。

本书适用于数学专业高年级本科生、工科研究生及高校教师使用。

图书在版编目(CIP)数据

矩阵分析/冯良贵,胡庆军编著 .—长沙:国防科技大学出版社,2010.6
ISBN 978 - 7 - 81099 - 767 - 6

Ⅰ.①矩… Ⅱ.①冯… ②胡… Ⅲ.①矩阵分析—研究生—教材 Ⅳ.①O151.21

中国版本图书馆 CIP 数据核字(2010)第 094500 号

国防科技大学出版社出版发行
电话:(0731)84572640 邮政编码:410073
http://www.gfkdcbs.com
责任编辑：耿 筠 责任校对：王 嘉
新华书店总店北京发行所经销
国防科技大学印刷厂印装
*
开本：787×1092 1/16 印张：10 字数：237 千
2010 年 6 月第 1 版第 1 次印刷 印数：1-1000 册
ISBN 978 - 7 - 81099 - 767 - 6
定价：29.00 元

前 言

本书在本科"线性代数"基础上，力求较全面系统地介绍矩阵分析的基本内容、方法及其在相关学科中的一些应用。取材较广，有一定的深度，编写过程中参考了众多国内外著名教材，竭心做到深入浅出，简明易懂。

全书共分七章，第一章旨在给出已学线性代数知识的必要回顾与提高；第二章主要讲述几种常见而又重要的矩阵分解，它们中有些在矩阵计算中发挥重要作用，有些则在理论证明中扮演重要角色；第三章介绍矩阵的广义逆理论，并着重说明其在求解相容和不相容线性方程组中的有关应用；第四章先介绍矩阵范数，然后讨论矩阵序列和矩阵级数的收敛性，作为对矩阵序列和矩阵级数的应用，给出了非负矩阵论中 Perron 定理的完整证明以及计算 A^+ 的几种迭代方法；第五章主要解决了针对一般的复函数 $f(z)$ 情形下，矩阵函数 $f(A)$ 的合理定义问题；第六章的主要目标是对元素为函数的矩阵建立起相应的微积分公式；最后在本书的第七章介绍特征值的估计分析，尤其着重论述关于特征值的著名戈氏圆盘定理及其推广形式。

众所周知，矩阵论在历史上至少可追溯到 Sylvester 与 Cayley，特别是 Cayley 的工作。在近代数学、工程技术、经济理论、管理科学、信息技术等领域中，大量地用到矩阵理论的知识和方法。这些理论和方法以其悠久的历史、丰富的内容和广泛的应用已自然地形成数学学科中一个重要分支——矩阵论。学习和掌握矩阵分析的基本理论与方法，对于工程技术人员、高等理工科院校研究生、本科生已成为必不可少的一项内容。矩阵分析已然是一门理论严谨、方法独特的数学基础课，它对培养学生的逻辑推理能力以及解决实际问题的能力等诸多方面具有极其重要的地位和作用。

感谢国防科技大学校、院、系各级领导的支持和关心，感谢新世纪优秀人才支持计划（NCET）及国防科技大学基础研究项目 JC08－02－03 的资助。由于时间仓促，错误在所难免，恳请同行专家和读者不吝指正。

作者

2009.10.20 于湖南·长沙

目 录

第 5 章　矩阵函数及其计算

第 6 章　函数矩阵的微积分及应用

第 7 章　特征值估计

第 1 章　线性变换的矩阵表示

本章旨在给出已学线性代数知识的必要回顾,主要内容包括对偶空间、多重线性型以及线性变换的矩阵表示等.本书除特别说明外,约定$\mathbb{F}$表示实数域$\mathbb{R}$或复数域$\mathbb{C}$,涉及的内积空间是指实内积空间或复内积空间.

1.1　对偶空间

域$\mathbb{F}$上的 n 维向量空间均同构于$\mathbb{F}^n$.设 V 为$\mathbb{F}$上向量空间,由 V 可产生一个新空间:$\mathrm{Hom}(V,\mathbb{F})$,记为 V^*,它由从 V 到$\mathbb{F}$的一切线性变换组成.一般地,人们称 V^* 为 V 的**一次对偶空间或对偶空间**.若$\{x_1,\cdots,x_n\}$为 V 的一组基,则$\{x_1^*,\cdots,x_n^*\}$便为 V^* 的一组基,称之为相应于$\{x_1,\cdots,x_n\}$的对偶基,其中 x_i^* 由下列条件给出:

$$x_i^*(x_j)=\delta_{ij}=\begin{cases}1 & j=i\\ 0 & j\neq i\end{cases}$$

明显地,下列映照为一个双线性函数:

$$[-,-]:V\times V^*\to\mathbb{F},\qquad (x,y)\to[x,y]=y(x)$$

关于双线性函数$[-,-]$,我们首先有:

命题 1.1.1　设 $\dim_{\mathbb{F}}V=n$,$\{x_1,\cdots,x_n\}$为 V 的一组基,$\{a_1,\cdots,a_n\}$为任意给定的 n 个纯量,则$\exists!\ y\in V^*$ s.t. $[x_i,y]=a_i,i=1,2,\cdots,n$.

证明:(1)取 $y=\sum\limits_{i=1}^{n}a_ix_i^*$,则有$[x_i,y]=a_i,i=1,2,\cdots,n$.

(2)若 $z\in V^*$ 也满足条件$[x_i,z]=a_i,i=1,2,\cdots,n$,则$[x_i,z-y]=0$,从而 $z=y$.

□

命题 1.1.2　设 $u,v\in{}_{\mathbb{F}}V,u\neq v,\dim V=n$,则$\exists\, y\in V^*$ s.t. $[u,y]\neq[v,y]$.特别地,对 V^n 中的任何非零元 x,$\exists\, y\in V^*$ s.t. $[x,y]\neq 0$.

证明:显然.

□

如前所述,从一个向量空间${}_{\mathbb{F}}V$出发,可构造$\mathbb{F}$上的向量空间 V^*;进而重复从 V 构造 V^* 的过程,又得向量空间${}_{\mathbb{F}}V^{**}$;注意到该过程总可以重复执行,因此,给定一个${}_{\mathbb{F}}V$,便得 $V^*,V^{**},V^{***},\cdots$,当$\dim_{\mathbb{F}}V=n$ 时,利用对偶基便知,$\dim V=\dim V^*=\dim V^{**}=\cdots$,由于

$$V^*=\mathrm{Hom}_{\mathbb{F}}(V,\mathbb{F});\ V^{**}=\mathrm{Hom}_{\mathbb{F}}(V^*,\mathbb{F})$$

因此,任给定 $x\in V$,赋值映照 $\phi_x=[x,-]:V^*\to\mathbb{F},y\to[x,y]$便为 V^{**} 中元.可直接验证知:映照 $\phi:V\to V^{**},x\to\phi_x=[x,-]$为单的线性映照.一般地,若向量空间 V(可能是有限维,也可能是无限维)满足 ϕ 为同构,则称 V 为一个自反空间.容易知道,当 V 为有限维向量空间时,V 总为一个自反空间.因此,对一个有限维向量空间而言,我们甚至可通过同构 ϕ 将 V 和 V^{**} 视为同一个空间.向量空间的子集的零化子空间是 V^* 的子空间,它具有以下性质:

命题 1.1.3 让 V 为向量空间,$\varnothing\neq S\subseteq V$,记 $S^0=\{y\in V^*\mid y(s)=0\text{ 对 }\forall s\in S\}$.若 $\dim V=n<\infty$,S 含有非零向量,则 $S^0\neq V^*$.

定理 1.1.1 设 $\dim V=n,M\leqslant V,\dim M=m$ 则 $M^0\leqslant V^*$ 且 $\dim M^0=n-m$.

证明: (1) $M^0\leqslant V^*$ 可通过直接验算可得.

(2)计算 M^0 的维数.

设 M 的一组基为 $x_1,\cdots,x_m$,扩充为 V 的一组基为:$x_1,\cdots,x_m;x_{m+1},\cdots,x_n$.于是 V^* 的一组基为 $x_1^*,\cdots,x_m^*;x_{m+1}^*,\cdots,x_n^*$.容易知道,$x_{m+1}^*,\cdots,x_n^*$ 均零化 M.又 $\forall y\in V^*$,y 可表为

$$y=a_1x_1^*+\cdots+a_mx_m^*+a_{m+1}x_{m+1}^*+\cdots+a_nx_n^*$$

于是 $y(M)=0\Rightarrow\begin{cases}a_1=0\\ \quad\vdots\\ a_m=0\end{cases}$,因此 $\operatorname{span}\{x_{m+1}^*,\cdots,x_n^*\}=M^0$.

□

推论 1.1.1 设 $\dim V=n,M\leqslant V$,则通过同构 ϕ 可将 M 看成 M^{00}.

证明: 首先$(M^0)^0$ 是 V^{**} 的子空间;其次,记 M 的一组基为 $x_1\cdots,x_m$,将之扩充为 V 的一组基:$x_1,\cdots,x_m;x_{m+1},\cdots,x_n$.于是 M^0 的一组基为:$x_{m+1}^*,\cdots,x_n^*$.V^{**} 的一组基为:$[x_1,-],\cdots,[x_m,-],[x_{m+1},-],\cdots,[x_n,-]$.具体验证可知:$[x_1,-],\cdots,[x_m,-]$即构成 M^{00}的一组基.因此,把$[x_1,-]$看成 $x_1,\cdots,[x_m,-]$看成 x_m 后,M 与 M^{00}便为同一空间.

□

定理 1.1.2 设向量空间 V(可能为无限维)有分解 $V=M\oplus N$.则有 $M^*\simeq N^0,N^*\simeq M^0,V^*=M^0\oplus N^0$.

证明: 事实上,$\forall y\in M^*$,$yp_M\in V^*$,其中 p_M 为 V 中沿 N 到 M 的射影,于是 $yp_M\in N^0$,因此有映射 $\varphi:M^*\to N^0,y\to yp_M$ 是同构.类似地可证:$N^*\simeq M^0$,进而 $V^*=M^0\oplus N^0$.

□

1.2 多重线性型

让我们先从双线性型开始.回顾:给定$_{\mathbb{F}}V$,$_{\mathbb{F}}W$ 为两线性空间,$\omega(-,-):V\times W\to\mathbb{F}$ 称为一个**双线性型**(或叫**双线性函数**)是指 $\omega(-,-)$对每个分量均是线性的.例如:

$[-,-]: V\times V^{*}\to\mathbb{F}$ 即为一个双线性型.易知,给定系数域$\mathbb{F}$上的两个线性空间 U 和 V,集合$\{\omega\mid\omega$ 为 $U\times V$ 上的双线性型$\}$在映照的通常加法及通常纯量乘法下仍构成向量空间.对此空间,我们有:

定理 1.2.1　设$_{\mathbb{F}}U$ 的一组基为$\{x_1,\cdots,x_m\}$,$_{\mathbb{F}}V$ 的一组基为$\{y_1,\cdots,y_n\}$,则

(1) $\forall A=(a_{ij})\in\mathbb{F}^{m\times n}$,$U\times V$ 上有且仅有一个双线性型 ω s.t. $(\omega(x_i,y_j))=(a_{ij})$.

(2) $U\times V$ 上的双线性型空间的一组基为$\{\omega_{pq}\mid p=1,2,\cdots,m;q=1,2,\cdots,n\}$,其中$\omega_{pq}$满足$(\omega_{pq}(x_i,y_j))=(\delta_{ip}\delta_{jq})$.因此$\dim(U\times V$ 上的双线性型空间$)=mn$.

证明: (1) $\forall x\in U$, $x=(a_1,\cdots,a_m)\begin{pmatrix}x_1\\ \vdots\\ x_m\end{pmatrix}$; $\forall y\in V$, $y=(b_1,\cdots,b_n)\begin{pmatrix}y_1\\ \vdots\\ y_n\end{pmatrix}$.定义 $\omega:(x,y)\to(a_1,\cdots,a_m)A\begin{pmatrix}b_1\\ \vdots\\ b_n\end{pmatrix}$,则有 ω 为一个双线性型,则$(\omega(x_i,y_j))=(a_{ij})$.另一方面,设双线性型 ω_1,ω_2 分别满足$(\omega_1(x_i,y_j))=(a_{ij})$,$(\omega_2(x_i,y_j))=(a_{ij})$,则有$((\omega_1-\omega_2)(x_i,y_j))=0$,于是$(\omega_1-\omega_2)(x,y)=0$,即 $\omega_1=\omega_2$.

(2)首先,下列线性映照为 $1-1$ 对应:$U\times V$ 上的双线性型空间$\to\mathbb{F}^{m\times n}$,$\omega\to(\omega(x_i,y_j))$.特别地,$\omega_{pq}\to E_{pq}$.因此$\{\omega_{pq}\}$为线性无关子集.进一步,$\forall\omega$ 为 $U\times V$ 上的双线性型,ω 可用ω_{pq}来线性表示.

□

定义 1.2.1　设 $V_1,\cdots,V_k$ 是 k 个$\mathbb{F}$上的向量空间,若映照 $\omega(x_1,\cdots,x_k): V_1\times\cdots\times V_k\to\mathbb{F}$ 对每个分量都是线性的,则 ω 叫做一个 **k-线性型**. $V_1\times\cdots\times V_k$ 上所有 k-线性型集合关于通常的函数加法及乘量乘法构成$\mathbb{F}$上的向量空间,称此空间为 $V_1\times\cdots\times V_k$ 的 **k-线性型空间**.当 $V_1=\cdots=V_k=V$ 时,我们称 $V_1\times\cdots\times V_k$ 上的 k-线性型空间为 V 上 k-线性型空间.设$\{x_{l_1}^{(1)}\}_{l_1\in L_1},\cdots,\{x_{l_k}^{(k)}\}_{l_k\in L_k}$分别为 $V_1,\cdots,V_k$ 的一组基,那么任何一个 $V_1\times\cdots\times V_k$ 上的 k-线性型 ω,它由这些$(x_{l_1}^{(1)},\cdots,x_{l_k}^{(k)})$所对应的值唯一确定.由此得:$\omega\xleftrightarrow{1-1\text{对应}}|L_1|\cdots|L_k|$维乘量组.因此 $V_1\times\cdots\times V_k$ 上的 k-线性型空间的维数为$(\dim V_1)\times\cdots\times(\dim V_k)$.特别地,$\{\omega_{(l_1,\cdots,l_k)}\}$为其一组基,其中 $\omega_{(l_1,\cdots,l_k)}$为满足下列条件的 k-线性型:

$$\omega_{(l_1,\cdots,l_k)}:(x_{j_1}^{(1)},\cdots,x_{j_k}^{(k)})\to\begin{cases}1, & \text{当}(j_1,\cdots,j_k)=(l_1,\cdots,l_k)\text{时}\\ 0, & \text{否则}\end{cases}$$

对 V 上 k-线性型空间,容易想到置换的技巧应该能用上.例如,让 S_k 表 k 元置换集,$\pi\in S_k$,ω 为 V 上的一个 k-线性型,令

$$\pi\omega:(x_1,\cdots,x_k)\to\omega(x_{\pi(1)},\cdots,x_{\pi(k)})$$

则有 $\pi\omega$ 仍为 V 上的一个 k-线性型.由此引出:

定义 1.2.2　设 ω 为 V 上的一个 k-线性型.若 $\forall\pi\in S_k$,均有 $\pi\omega=\omega$.则称 ω 为一个

对称 k - 线性型.

V 上的所有对称 k - 线性型仍构成向量空间,它是 V 上 k - 线性型空间的子空间.从任何给定的 V 上 k - 线性型 ω 出发,下列 k - 线性型总为一个对称 k - 线性型: $\sum\limits_{\pi\in S_k}\pi\omega$.

在多重线性代数中,下列两种类型的 k - 线性型也是十分重要的.

(1)斜对称型

设 ω 为 V 上的一个 k - 线性型.若对任何奇置换 π,均有:$\pi\omega=-\omega$,则称 ω 为**斜对称的**;任何给定的 k - 线性型 ω, $\sum\limits_{\pi\in S_n}(\mathrm{sgn}\pi)\pi\omega$ 即为一个斜对称 k - 线性型,其中 $\mathrm{sgn}\pi$ 表示置换 π 的符号,即

$$\forall\,\pi\in S_k,\mathrm{sgn}\pi=\begin{cases}1, & \pi\text{ 为偶置换}\\ -1, & \pi\text{ 为奇置换}\end{cases}$$

(2)交错型

设 ω 为 V 上的 k - 线性型.若 $\omega(x_1,\cdots,x_k)$满足:一旦有两个分量取值相等,则 $\omega(x_1,\cdots,x_k)$便取值为 0.此时称 ω 为一个**交错 k - 线性型**.

交错型与斜对称型有如下关系:每个交错的 k - 线性型必为斜对称的;反之,当线性空间 V 的系数域 $\mathbb{F}$ 的特征不为 2 时,每个斜对称的 k - 线性型也必为交错 k - 线性型.

下述定理表明:对一个 n 维向量空间 V 中的 n 个向量 $x_1,\cdots,x_n$, $\{x_1,\cdots,x_n\}$的线性相关性可借助于 V 上的任何一个非零交错 n - 线性型 ω 来判定.

定理 1.2.2 设 $\dim_{\mathbb{F}}V=n$, $x_1,\cdots,x_n$ 为 V 中的 n 个向量,ω 为一个 V 上的非零交错 n - 线性型,则:

(1) $\{x_1,\cdots,x_n\}$为 $L.\mathrm{ind}.\Leftrightarrow\omega(x_1,\cdots,x_n)\neq0$.

(2)若 ω' 也为 V 上的一个非零交错 n - 线性型,则 ω 与 ω' 必线性相关.

证明: (1)$\Rightarrow$ $\{x_1,\cdots,x_n\}$为 $L.\mathrm{ind}.$,则$\{x_1,\cdots,x_n\}$构成 V 的一组基.若 $\omega(x_1,\cdots,x_n)=0$,则$\forall(y_1,\cdots,y_n)\in\underbrace{V\times\cdots\times V}_{n}$,将 y_i 写成 $x_1,\cdots,x_n$ 的线性组合,展开 $\omega(y_1,\cdots,y_n)$后即得到 0,这导致 $\omega=0$,矛盾.

$\Leftarrow$ 若 $x_1,\cdots,x_n$ 线性相关,则不妨说 $x_1=a_2x_2+\cdots+a_nx_n$,于是 $\omega(a_2x_2+\cdots+a_nx_n,x_2,\cdots,x_n)=0$,矛盾于 $\omega(x_1,\cdots,x_n)\neq0$.

(2)取$\{e_1,\cdots,e_n\}$为 V 的一组基.$\forall(y_1,\cdots,y_n)\in\underbrace{V\times\cdots\times V}_{n}$,有

$$(y_1,\cdots,y_n)=(e_1,\cdots,e_n)\begin{pmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nn}\end{pmatrix}$$

$\omega(y_1,\cdots,y_n)=\omega(e_1,\cdots,e_n)\times$某个$\mathbb{F}$ 的元 c,该 c 仅由$\begin{pmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{n1} & \cdots & a_{nn}\end{pmatrix}$决定;同样地,$\omega'(y_1,\cdots,y_n)=\omega'(e_1,\cdots,e_n)\times$某个$\mathbb{F}$ 的元 c'.注意到 $c=c'$,而 $\omega(e_1,\cdots,e_n)$与 $\omega'(e_1,$

$\cdots,e_n$)显然是线性相关的.因此,ω 与 ω' 线性相关.

□

定理 1.2.2 中的(2)表明,对 n 维向量空间 V 而言,V 上的交错 n－线性型构成的空间的维数$\leqslant 1$.进一步,我们还有:

定理 1.2.3　设 $\dim_{\mathbb{F}}V=n$.则

(1)当 $k>n$ 时,V 上的交错 k－线性型只有 1 个,为 0;

(2)V 上的交错 n－线性型构成的空间是 1 维的.

证明:(1)当 $k>n$ 时,任何 k 个向量 $x_1,\cdots,x_k$ 均是线性相关,从而 V 上的交错 k－线性型均为 0.

(2)当 $k=n$ 时,只需证明 V 上存在一个非零的交错 k－线性型即可.使用归纳法.当 $n=1$ 时,显然存在非零交错 1－线性型(事实上,V^* 的任何一个非零元即是一例).

设 $n=l$ 时,存在一个非零交错 l－线性型 v.则对 $n=l+1$ 时,先对 V 的一个 l 维子空间 W 取其一个非零交错 l－线性型 v,将 $v(p_W(x_1),\cdots,p_W(x_l))$仍记为 v,这里 $x_1,\cdots,x_l\in V$,p_W 表示 V 到 W 的一个投射,此时 v 便为 V 上的一个交错 l－线性型.于是$\exists$向量 $x_1^0,\cdots,x_l^0\in W\subseteq V$ 使得 $v(x_1^0,\cdots,x_l^0)\neq 0$.再取 $x_{l+1}^0\notin \mathrm{span}\{x_1^0,\cdots,x_l^0\}$,于是空间$\mathbb{F}x_{l+1}^0+\mathrm{span}\{x_1^0,\cdots,x_l^0\}=\mathbb{F}x_{l+1}^0\oplus\mathrm{span}\{x_1^0,\cdots,x_l^0\}$即为空间 V,其中$\{x_1^0,\cdots,x_l^0,x_{l+1}^0\}$构成 V 的一组基,这里利用了定理 1.2.2 中的(1).

在 V 的对偶基$(x_1^0)^*,\cdots,(x_{l+1}^0)^*$中,令 $u=(x_{l+1}^0)^*$,则 u 为 V 到 $\mathbb{F}$ 的线性函数且满足

$$\begin{cases}u(x_1^0)=\cdots=u(x_l^0)=0\\ u(x_{l+1}^0)\neq 0\end{cases}$$

构造 $\omega(x_1,\cdots,x_l,x_{l+1})$如下:

$$\omega(x_1,\cdots,x_l,x_{l+1})=\left[\sum_{i=1}^{l}(i,l+1)(v(x_1,\cdots,x_l)u(x_{l+1}))\right]-v(x_1,\cdots,x_l)u(x_{l+1})$$

其中 $v(x_1,\cdots,x_l)$表示 $v(p_w(x_1),\cdots,p_w(x_l))$,$x_1,\cdots,x_l,x_{l+1}\in V$,这里$(i,l+1)$表示对 i 与 $l+1$ 进行互换,则容易验证:

①$\omega\neq 0$,且 ω 为$(l+1)$－线性型;

②设 $x_i=x_j$,则

$$\begin{aligned}&\omega(x_1,\cdots,x_i,\cdots,x_j,\cdots,x_{l+1})\\&=v(x_{l+1},x_2,\cdots,x_l)u(x_1)+\cdots+v(x_1,\cdots,x_{l-1},x_{l+1})u(x_l)-v(x_1,\cdots,x_l)u(x_{l+1})\\&=\begin{cases}0 & i,j<l+1\\ 0 & i,j\text{ 中有一个为 }l+1\text{ 时}\end{cases}\end{aligned}$$

因此 ω 为交错型.

□

1.3 线性变换的表示阵

对线性空间$_{\mathbb{F}}V$和$_{\mathbb{F}}W$,选定 V 的一组基$\{\xi_1,\cdots,\xi_m\}$和 W 的一组基$\{\eta_1,\cdots,\eta_n\}$,则映照 $\phi: L(V,W)\to\mathbb{F}^{n\times m}$($\phi(\sigma)=[\sigma]$,其中$[\sigma]$由式子 $\sigma(\xi_1,\cdots,\xi_m)=(\eta_1,\cdots,\eta_n)[\sigma]$唯一决定)建立了 $L(V,W)$到$\mathbb{F}^{n\times m}$之间的一一对应,通常称矩阵$[\sigma]$为 σ 在基$\{\xi_1,\cdots,\xi_m\}$和$\{\eta_1,\cdots,\eta_n\}$下的**表示矩阵**.易知只要 $\sigma\tau$ 有定义,就有$[\sigma\tau]=[\sigma][\tau]$.借助于线性变换的表示阵,我们可将矩阵看成线性变换,同时又可通过研究矩阵来研究线性变换.因此,可以说矩阵理论也就是向量空间的线性变换理论.

下面几个结论是显然的.

(1)设 $\dim_{\mathbb{F}}V=n$,$\sigma\in L(V,V)$,则 σ 为同构 $\Leftrightarrow$ σ 为单射 $\Leftrightarrow$ σ 为满射.

(2)$\forall\sigma\in L(V,V)$,若$\{\xi_1,\cdots,\xi_m\}$,$\{\eta_1,\cdots,\eta_m\}$为 V 的两组基,则 σ 在$\{\xi_1,\cdots,\xi_m\}$下的表示阵与 σ 在$\{\eta_1,\cdots,\eta_m\}$下的表示阵是相似矩阵,这里 σ 在$\{\xi_1,\cdots,\xi_m\}$下的表示阵是指通过 $\sigma(\xi_1,\cdots,\xi_m)=(\xi_1,\cdots,\xi_m)M$ 所确定的矩阵 M.

(3)设 $\sigma\in L(V,V)$,$\dim_{\mathbb{F}}V=n$.若 W 是 V 的不变子空间,则在适当选取 V 的基下,σ 的表示阵可为$\begin{pmatrix}A_1 & A_2\\ 0 & A_4\end{pmatrix}$的形式,其中 A_1,A_4 均为方阵.

(4)n 阶方阵A 可相似于对角块形式$\begin{pmatrix}M & \\ & N\end{pmatrix}$当且仅当将$A$ 看成n 维向量空间 V 上的线性变换时有不变空间 W_1 和 W_2 满足 $V=W_1\oplus W_2$.

有时,将矩阵看成线性变换后能有助于较方便地获得有关矩阵的相关结果,如下述定理中的(3).

定理 1.3.1 设 $A\in\mathbb{C}^{n\times n}$,则以下几条成立.

(1)若秩 $A=\rho>1$,则 A 可解为$A=A_1+\cdots+A_\rho$,其中秩 $A_i=1(i=1,2,\cdots,\rho)$.

(2)秩 $A\leqslant 1$ 等价于 A,可写成 $A=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}(\alpha_1,\cdots,\alpha_n)$,其中 β_i 与 α_i 均属于$\mathbb{C}(i=1,2,\cdots,n)$.

(3)$\exists$可逆的 $P\in\mathbb{C}^{n\times n}$,使得$(PA)^2=PA$.

证明:让$\{\xi_1,\cdots,\xi_n\}$为$_{\mathbb{C}}V$ 的一组基,将 A 看成下列线性变换:

$$\sigma: V\to V$$
$$\sigma(\xi_1,\cdots,\xi_n)=(\xi_1,\cdots,\xi_n)A$$

(1)秩 $A=\rho\Leftrightarrow\dim\sigma(V)=\rho\Rightarrow\sigma(V)=\mathrm{span}\{\sigma x_1,\cdots,\sigma x_\rho\}$,其中$\{\sigma x_1,\cdots,\sigma x_\rho\}$为 $\mathrm{Im}\,\sigma$ 的一组基,结果,$\sigma(V)=\mathbb{C}\,\sigma x_1\oplus\cdots\oplus\mathbb{C}\,\sigma x_\rho$.考察映照:$V\xrightarrow{\sigma}\sigma(V)\xrightarrow{\pi_i}\mathbb{C}\,\sigma x_i\xrightarrow{\lambda_i}V$,其中

π_i 为投射,λ_i 为嵌入.则 $\lambda_i\pi_i\sigma\in L(V,V)$, $\sigma=\sum_{i=1}^{\rho}\lambda_i\pi_i\sigma$,注意到 $\lambda_i\pi_i\sigma$ 是秩为 1 的线性变换,因此 $A=\sum_{i=1}^{\rho}[\lambda_i\pi_i\sigma]$,如所愿.

(2)⇒　当秩 $A=0$ 时,$A=\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix}(0,\cdots,0)$.当秩 $A=1$ 时,A 有分解,$A=P\begin{pmatrix}1&0\\0&0\end{pmatrix}Q$,其中 P、Q 可逆.注意到 $P\begin{pmatrix}1&0\\0&0\end{pmatrix}=\begin{pmatrix}p_{11}&0\\ \vdots&\vdots\\ p_{n1}&0\end{pmatrix}$,$\begin{pmatrix}1&0\\0&0\end{pmatrix}Q=\begin{pmatrix}q_{11}&\cdots&q_{1n}\\0&\cdots&0\end{pmatrix}$.故

$$A=\begin{pmatrix}p_{11}\\ \vdots\\ p_{n1}\end{pmatrix}(q_{11}\quad\cdots\quad q_{1n})$$

⇐　显然.

(3)让 R 与 N 分别表示 σ 的象空间与零空间,取 R 的基为 $\{x_1,\cdots,x_\rho\}$,将其扩充为 V 的基 $\{x_1,\cdots,x_\rho;x_{\rho+1},\cdots,x_n\}$.取 $\{y_1,\cdots,y_\rho\}$ 满足 $\begin{cases}\sigma(y_1)=x_1\\ \quad\vdots\\ \sigma(y_\rho)=x_\rho\end{cases}$,则 $y_1,\cdots,y_\rho$ 线性无关,又将 $\{y_1,\cdots,y_\rho\}$ 并上 N 的一组基 $\{y_{\rho+1},\cdots,y_n\}$ 得 V 的基 $\{y_1,\cdots,y_\rho;y_{\rho+1},\cdots,y_n\}$.令线性变换 τ 为

$$\tau:V\to V,\tau(x_i)=y_i(i=1,2,\cdots,n)$$

于是 τ 为可逆的,且

$$\tau\sigma(y_i)=y_i(i=1,2,\cdots,\rho),\tau\sigma(y_j)=0\ (j=\rho+1,\rho+2,\cdots,n)$$

因此 $(\tau\sigma)^2=\tau\sigma$.令 τ 在基 $\{\xi_1,\cdots,\xi_n\}$ 下的表示阵为 P,则有 $(PA)^2=PA$.

□

定义 1.3.1　设 $\mathbb{A}\in L(V,V)$,其中 V 为域 $\mathbb{F}$ 上的有限维向量空间,线性变换:

$$\mathbb{A}':V^*\to V^*$$
$$y\to[\mathbb{A}-,y]=y(\mathbb{A}-)$$

称为 $\mathbb{A}$ 的**伴随变换**或对偶,有时也简称为 $\mathbb{A}$ 的**伴随**.

显见,$\mathbb{A}$ 的伴随 $\mathbb{A}'$ 有如下性质:$\forall x\in V,\forall y\in V^*$,$[\mathbb{A}x,y]=[x,\mathbb{A}'y]$,因此:$\mathbb{A}$ 与 $\mathbb{A}'$ 之间可由下列式子将它们联系起来:

$$[\mathbb{A}-,-]=[-,\mathbb{A}'-]$$

进一步,映射

$$\psi:L(V,V)\to L(V^*,V^*),\mathbb{A}\to\mathbb{A}'$$

实际上建立了 $L(V,V)$ 到 $L(V^*,V^*)$ 的一个反代数同构.

现假定 $\{\xi_1,\cdots,\xi_n\}$ 为 V 的一组基,$\{\xi_1^*,\cdots,\xi_n^*\}$ 为其对偶基.由前面已知,$\mathbb{A}$ 在 $\{\xi_1,\cdots,\xi_n\}$ 下有表示阵 $A=(a_{ij})$.现在来考察 $\mathbb{A}'$ 在其基 $\{\xi_1^*,\cdots,\xi_n^*\}$ 下的表示阵 $[\mathbb{A}']$.注意到

有等式：

$$\mathscr{A}'(\xi_1^*,\cdots,\xi_n^*)=(\xi_1^*,\cdots,\xi_n^*)[\mathscr{A}']$$

因此$[\mathscr{A}']$的(i,j)位置元素为$[\xi_i,\mathscr{A}'\xi_j^*]$.

但

$$\begin{aligned}[\xi_i,\mathscr{A}'\xi_j^*]&=[\mathscr{A}\xi_i,\xi_j^*]\\&=[a_{1i}\xi_1+\cdots+a_{ni}\xi_n,\xi_j^*]\\&=a_{ji}\end{aligned}$$

因此$[\mathscr{A}']=A^{\mathrm{T}}$($A$ 的转置).

下面进一步给出几类重要的线性变换的表示矩阵.

(Ⅰ)射影的表示阵

设 $V=M\oplus N$, p_M 为 V 到 M 的投射, $P_M:V\xrightarrow{p_M}M\longrightarrow V$ 称为沿 N 到 M 上的**射影**. 对于一个给定的 $E\in L(V,V)$,容易验证:E 为某子空间上的射影 $\Leftrightarrow E^2=E\Leftrightarrow(1-E)^2=(1-E)$.进一步,若 E 是沿 N 到 M 上的射影,则 $M=\{x\in V\mid Ex=x\}$, $N=\{x\in V\mid Ex=0\}=\ker E$,而 $1-E$ 为沿 M 到 N 上的射影.因此,任何一个射影的表示阵即为幂等阵;反之任何一个幂等阵也可看成一个射影.

(Ⅱ)内积空间中共轭伴随的表示阵

所谓内积空间是指赋予了一个内积$\langle\cdot\rangle$的向量空间.对一个有限维向量空间 V 而言,设 M 为 V 的子空间,通常满足 $M\oplus N=V$ 的子空间 N 未必唯一.但 V 为有限维内积空间时,满足 $M\oplus N=V$ 且 $N\perp M$ 的 N 却只有 1 个,它就是 M 的正交补 $M^{\perp}$.现设$(V,\langle\cdot\rangle)$为 n 维内积空间,$\mathscr{A}\in L(V,V)$.通过令

$$\langle\mathscr{A}x,y\rangle=\langle x,\mathscr{A}^*y\rangle,\text{对每个 } x\in V \text{ 成立}$$

来定义一个 V 到 V 的线性变换 $\mathscr{A}^*$,注意:这里用到了 Riese 表示定理,从而使定义合理.该$\mathscr{A}^*$通常称为$\mathscr{A}$ 的共轭伴随.若$\mathscr{A}$ 在标准正交基$\{e_1,\cdots,e_n\}$下的表示阵为 A,那么$\mathscr{A}^*$ 在$\{e_1,\cdots,e_n\}$下的表示阵$[\mathscr{A}^*]$会是怎样呢?

事实上,$\mathscr{A}^*(e_1,\cdots,e_n)=(e_1,\cdots,e_n)[\mathscr{A}^*]\Rightarrow[\mathscr{A}^*]$的$(i,j)$位置元素为$\langle\mathscr{A}^*e_j,e_i\rangle$.

但

$$\begin{aligned}\langle\mathscr{A}^*e_j,e_i\rangle&=\overline{\langle e_i,\mathscr{A}^*e_j\rangle}\\&=\overline{\langle\mathscr{A}e_i,e_j\rangle}\\&=\overline{\langle a_{1i}e_1+\cdots+a_{ni}e_n,e_j\rangle}\\&=\overline{a_{ji}}\end{aligned}$$

因此$[\mathscr{A}^*]=A^*$(转置共轭).

对内积空间 V 上的一个线性变换E,若$\exists M\leqslant V$, s.t. $V=M\oplus M^{\perp}$,且 $E=P_M$,此时我们就称 E 为M 上的一个垂直射影.直接验证可知,下列命题成立.

命题 1.3.1 设 E 为有限维内积之间 V 上的线性变换,则以下等价:

(1)E 为垂直射影.

(2)$E=E^2=E^*$;(2)$'E$ 在 V 的任何一组标正基下的表示阵均为自共轭幂等阵.

(3) $E=E^2$，对 $\forall x\in V$，且 $\|Ex\|\leqslant\|x\|$.

证明： (1)⇒(2)　$E=P_M$，其中 M 满足 $V=M\oplus M^{\perp}$，$V\xrightarrow{P_M}M\rightarrow V$. E 的幂等性是显然的. 下证 $E=E^*$. 取 M 的一组标正基拼上 $M^{\perp}$ 的一组标正基得 V 的一组标正基 $\mathbb{B}$. 则 E 在 $\mathbb{B}$ 下的表示阵为 $\begin{pmatrix} I_r & \\ & 0\end{pmatrix}$，进而 E^* 在 $\mathbb{B}$ 下的表示阵为 $\begin{pmatrix} I_r & \\ & 0\end{pmatrix}^*=\begin{pmatrix} I_r & \\ & 0\end{pmatrix}$，从而 $E=E^*$.

(2)⇒(1)　首先由 $E=E^2$ 知 E 为射影. 于是 $V=\mathrm{Im}E\oplus\ker E$. $\forall x\in\ker E$，$\forall y\in\mathrm{Im}E$，y 可表为 $y=Ez$. 于是

$$\begin{aligned}\langle x,y\rangle&=\langle x,Ez\rangle\\&=\langle x,E^*z\rangle\\&=\langle Ex,z\rangle\\&=\langle 0,z\rangle=0\end{aligned}$$

这表明 $\ker E\perp\mathrm{Im}E$. 因此 E 为垂直射影.

(1)⇒(3)　$\forall x\in V$，有

$$\begin{aligned}\|Ex\|^2&=\langle Ex,Ex\rangle\\&=\langle x,E^*Ex\rangle\\&=\langle x,Ex\rangle\\&\leqslant\|x\|^2\|Ex\|^2\end{aligned}$$

结果 $\|Ex\|\leqslant\|x\|$.

(3)⇒(1)　由 $E=E^2$ 得 $V=\mathrm{Im}E\oplus\ker E$. 下证 $\mathrm{Im}E\perp\ker E$ 即可. 为此，将 V 写为 $V=(\ker E)^{\perp}\oplus\ker E$. 于是 $\forall\alpha\in(\ker E)^{\perp}$，$E\alpha-\alpha\in\ker E$，进而 $E\alpha$ 可写为 $E\alpha=\alpha+y$，其中 $y\in\ker E$ 且 $\alpha\perp y$. 结果

$$\|\alpha\|^2\geqslant\|E\alpha\|^2=\|\alpha\|^2+\|y\|^2$$

导致 $y=0$，$(\ker E)^{\perp}\subseteq\mathrm{Im}E$. 由 $\dim\mathrm{Im}E=\dim(\ker E)^{\perp}$ 立即为：$\mathrm{Im}E=(\ker E)^{\perp}$.

□

关于幂等阵的组合，给出如下两个命题.

命题 1.3.2　设 $\mathbb{F}$ 表示数域. $A,B\in\mathbb{F}^{n\times n}$，$A^2=A$，$B^2=B$，则

(1) $A+B$ 仍幂等 ⇔ $AB=BA=0$. 进一步 $A+B$ 仍幂等时，$\ker(A+B)=\ker A\cap\ker B$，$\mathrm{Im}(A+B)=\mathrm{Im}A\oplus\mathrm{Im}B$.

(2) $A-B$ 仍幂等 ⇔ $AB=BA=B$. 进一步 $A-B$ 仍幂等时，$\ker(A-B)=\ker A\oplus\mathrm{Im}B$，$\mathrm{Im}(A-B)=\mathrm{Im}A\cap\ker B$.

(3) 若 $C=AB=BA$，则 C 仍幂等且 $\ker C=\ker A+\ker B$，$\mathrm{Im}C=\mathrm{Im}A\cap\mathrm{Im}B$.

证明： (1) $(A+B)^2=A+B\Rightarrow A^2+AB+BA+B^2=A+B\Rightarrow AB=-BA$. 因此有 $AB=AAB=-ABA$ 及 $ABA=-BAA=-BA$ 同时成立. 这表明 $AB=BA$，故此时 $AB=BA=0$. 反之显然. 当 $A+B$ 幂等时，有

$$\ker(A+B)=\{x\in\mathbb{F}^{n\times 1}\mid(A+B)x=0\}$$

$$= \{x \in \mathbb{F}^{n\times 1} \mid Ax = -Bx\}$$

由 $Ax = -Bx$ 导致 $\begin{cases} A^2x = -ABx = 0 \\ B^2x = 0 \end{cases}$ 即 $\begin{cases} Ax = 0 \\ Bx = 0 \end{cases}$. 因此 $\ker(A+B) = \ker A \cap \ker B$. 易知此时 $\mathrm{Im}(A+B) = \mathrm{Im}A \oplus \mathrm{Im}B$.

(2)考虑 $I-(A-B)$即可.

(3) $\forall x \in \ker C$, $ABx = 0 \Rightarrow Bx \in \ker A$. 注意到 $x = Bx + (I-B)x$, 因此 $\ker C \subseteq \ker A + \ker B$, $\ker A + \ker B \subseteq \ker C$ 是显然的, 因此 $\ker C = \ker A + \ker B$. 易知 $\mathrm{Im}C = \mathrm{Im}A \cap \mathrm{Im}B$.

□

命题 1.3.3 设 $A_1, \cdots, A_n$ 均为自共轭幂等 n 阶方阵, 则 $A_1 + \cdots + A_n$ 仍为幂等阵 $\Leftrightarrow$ $A_iA_j = 0$ 对任意的 $i \neq j$.

证明: $\Rightarrow$ 将 $A_1 + \cdots + A_n$ 分别看成垂直射影, 且假定 $A = A_1 + \cdots + A_n$ 仍是垂直射影. $\forall x \in \mathrm{Im}A_i$, 有

$$\begin{aligned} \|x\|^2 \geqslant \|Ax\|^2 &= \langle Ax, Ax\rangle = \langle A^*Ax, x\rangle \\ &= \langle \sum_j A_jx, x\rangle = \sum_j \langle A_j^2x, x\rangle \\ &= \sum_j \|A_jx\|^2 \geqslant \|A_ix\|^2 \\ &= \|x\|^2 \end{aligned}$$

结果 $A_jx = 0$($j \neq i$ 时), 也就是 $A_jA_i = 0$ 对 $i \neq j$.

$\Leftarrow$ 显然.

□

命题 1.3.4 设 A, B 均为自共轭阵, 记号 $A \leqslant B$ 表示 $B - A$ 为半正定阵. 若 A, B 均为自共轭幂等阵, 则有以下几条等价:

(1) $A \leqslant B$;

(2) $\forall x \in \mathbb{C}^{n\times 1}$ 有 $\|Ax\| \leqslant \|Bx\|$;

(3) $\mathrm{Im}A \subseteq \mathrm{Im}B$;

(4) $BA = A$;

(5) $AB = A$.

证明: (1)$\Rightarrow$(2) $0 \leqslant \langle (B-A)x, x\rangle = \langle Bx, x\rangle - \langle Ax, x\rangle = \|Bx\|^2 - \|Ax\|^2$.

(2)$\Rightarrow$(3) 任取 $x \in \mathrm{Im}A$, $\|x\| = \|Ax\| \leqslant \|Bx\| \leqslant \|x\|$, 从而 $\|Bx\| = \|x\|$, 再由 B 为自共轭幂等导致 $x \in \mathrm{Im}B$.

(3)$\Rightarrow$(4) 显然.

(4)$\Leftrightarrow$(5) 两边取转置共轭即可.

(4)$\Rightarrow$(1) $\langle (B-A)x, x\rangle = \langle B(I-A)x, x\rangle$. 由 $B(I-A) = (I-A)B$ 知 $B(I-A)$仍是幂等自共轭阵, 故$\langle (B-A)x, x\rangle = \langle B(I-A)x, B(I-A)x\rangle \geqslant 0$, 即有 $B \geqslant A$.

□

(Ⅲ)保距变换的表示阵

对内积空间$(V,\langle\cdot\rangle)$,有一种特殊的线性变换——**保距变换**,它被定义为:$\forall x,y\in V$均有$\| Ux-Uy\|=\|x-y\|$的线性变换U,这里$\|\cdot\|$为通常的欧几里得范数.由于下列几条等价:

(1)U为保距变换;

(2)$U^*U=1$;

(3)$\langle Ux,Uy\rangle=\langle x,y\rangle,\forall x,y\in V$;

(4)$\|Ux\|=\|x\|,\forall x\in V$;

(5)U将标正基变换为标正基.

因此,当内积空间为欧氏空间时,U在任何一组标正基下的表示阵为正交阵;当内积空间为酉空间时,U在任何标正基下的表示阵为酉矩阵.下面给出正交阵的标准型,关于酉矩阵标准型的相关结果在下一章的定理2.3.3中给出.为了给出实正交阵的标准型,先来介绍实向量空间的复化空间的有关知识.

让V为实向量空间,定义集合V^+为

$$V^+=V+\mathrm{i}V=\{\alpha+\mathrm{i}\beta\mid\alpha,\beta\in V,\mathrm{i}\text{ 为未定元},\mathrm{i}^2=-1\}$$

进一步定义复数与V^+中元之间的纯量乘法如下:

$$(a+b\mathrm{i})(\alpha+\mathrm{i}\beta)=(a\alpha-b\beta)+\mathrm{i}(b\alpha+a\beta),\text{对}\forall a+b\mathrm{i}\in\mathbb{C}\text{ 和}\forall\alpha+\mathrm{i}\beta\in V^+$$

容易知道,此V^+构成$\mathbb{C}$上的一个向量空间,该空间便称为${}_{\mathbb{R}}V$的**复化空间**.${}_{\mathbb{R}}V$的复化空间${}_{\mathbb{C}}V^+$有如下性质:

(1)${}_{\mathbb{R}}V^+={}_{\mathbb{R}}V\oplus_{\mathbb{R}}(\mathrm{i}V)$;

(2)若$\{\xi_1,\cdots,\xi_n\}$为${}_{\mathbb{R}}V$中的一个线性无关向量集,则$\{\xi_1,\cdots,\xi_n\}$在${}_{\mathbb{C}}V^+$中也为关于$\mathbb{C}$的线性无关向量集,因此不难发现$\dim_{\mathbb{R}}V=\dim_{\mathbb{C}}V^+$.

现将${}_{\mathbb{R}}V$中的每个线性变换$\mathbb{A}$,按如下方式扩展为${}_{\mathbb{C}}V^+$中的线性变换$\mathbb{A}^+$:

$$\mathbb{A}^+:{}_{\mathbb{C}}V^+\to{}_{\mathbb{C}}V^+$$
$$\alpha+\mathrm{i}\beta\to\mathbb{A}\alpha+\mathrm{i}\,\mathbb{A}\beta$$

则有映照$\Gamma:L({}_{\mathbb{R}}V,{}_{\mathbb{R}}V)\to L({}_{\mathbb{C}}V^+,{}_{\mathbb{C}}V^+)(\mathbb{A}\to\mathbb{A}^+)$保持加法、乘法以及关于实数的纯量乘法.倘若$\mathbb{A}$在${}_{\mathbb{R}}V$的某基$\{\xi_1,\cdots,\xi_n\}$下的表示阵为$A$,则有$\mathbb{A}^+$在$\{\xi_1,\cdots,\xi_n\}$下的表示阵也为$A$.对实向量空间进行复化,一个直接的应用是,我们能轻易地证得如下结论:

定理1.3.2　n维实向量空间V中的每一个线性变换$\mathbb{A}$,都有一个维数等于1或2的子空间作为$\mathbb{A}$的不变子空间.

证明: 设$\mathbb{A}^+$的一个特征值为$x+\mathrm{i}y$,相应于此特征值的一个特征向量是$\alpha+\mathrm{i}\beta$,则由

$$\begin{aligned}\mathbb{A}\alpha+\mathrm{i}\,\mathbb{A}\beta&=\mathbb{A}^+(\alpha+\mathrm{i}\beta)=(x+\mathrm{i}y)(\alpha+\mathrm{i}\beta)\\&=(x\alpha-y\beta)+\mathrm{i}(x\beta+y\alpha)\end{aligned}$$

得:

$$\begin{cases}\mathbb{A}\alpha=x\alpha-y\beta\\\mathbb{A}\beta=x\beta+y\alpha\end{cases}$$

因此,$\mathrm{span}\{\alpha,\beta\}$作为${}_{\mathbb{R}}V$的子空间是$\mathbb{A}$的不变子空间.

□

当被复化的实向量空间为欧氏空间$\langle V,\langle\cdot\rangle\rangle$时，$_{\mathbb{C}}V^{+}$还可通过引进下列内积，从而使之成为酉空间：$\langle\alpha_1+\mathrm{i}\beta_1,\alpha_2+\mathrm{i}\beta_2\rangle=\langle\alpha_1,\alpha_2\rangle+\langle\beta_1,\beta_2\rangle+\mathrm{i}(\langle\beta_1,\alpha_2\rangle-\langle\alpha_1,\beta_2\rangle)$，其中$\alpha_1,\beta_1,\alpha_2,\beta_2\in V$.此时，$\|\alpha+\mathrm{i}\beta\|^2=\|\alpha\|^2+\|\beta\|^2$对任何$\alpha,\beta\in V$皆成立.

有了以上知识，现着手研究正交矩阵的标准型问题.设U为n阶实正交阵.我们考察n维实内积空间V相应于U的正交变换$\mathbb{U}$.将V复化为酉空间$_{\mathbb{C}}V^{+}=V+\mathrm{i}V$，同时也将$\mathbb{U}$复化为$\mathbb{U}^{+}$，则有$\mathbb{U}^{+}$仍为酉变换.稍作推导即知，$\mathbb{U}^{+}$还有以下性质：

(1)$\mathbb{U}^{+}$的非实特征值必共轭成对出现.若用$(V^{+})_\lambda$表示$\mathbb{U}^{+}$关于λ的特征子空间，则映照$\psi:(V^{+})_\lambda\to(V^{+})_{\bar\lambda},v\to\bar v$为一一对应.

(2)$\mathbb{U}^{+}$的酉性保证了$\mathbb{U}^{+}$的任两个互异特征子空间相互垂直.进一步，当特征值λ不为实数时，若$\alpha+\mathrm{i}\beta\in(V^{+})_\lambda$，则由$\overline{\alpha+\mathrm{i}\beta}=\alpha-\mathrm{i}\beta\in(V^{+})_{\bar\lambda}$及$\langle\alpha+\mathrm{i}\beta,\overline{\alpha+\mathrm{i}\beta}\rangle=0$得：

$$\begin{cases}\alpha\perp\beta\\ \|\alpha\|=\|\beta\|=\dfrac{1}{\sqrt2}\|\alpha+\mathrm{i}\beta\|\end{cases}$$

因此，若$\{\eta_1=\alpha_1+\mathrm{i}\beta_1,\cdots,\eta_k=\alpha_k+\mathrm{i}\beta_k\}$为$(V^{+})_\lambda$的标正基，则$\{\bar\eta_1,\cdots,\bar\eta_k\}$为$(V^{+})_{\bar\lambda}$的标正基，进而$\{\sqrt2\alpha_1,\cdots,\sqrt2\alpha_k\}\cup\{\sqrt2\beta_1,\cdots,\sqrt2\beta_k\}$为$V$中一个标准正交系.

(3)$\mathbb{U}^{+}$的特征值均坐落在单位圆周上.若用V_1表示$\mathbb{U}$关于1的V的特征子空间及Y_1为V_1的一组标正基；V_{-1}表示$\mathbb{U}$关于-1的V的特征子空间及Y_{-1}为V_{-1}的一组标正基，则由于

$$\begin{cases}(V^{+})_1=V_1+\mathrm{i}V_1\\(V^{+})_{-1}=V_{-1}+\mathrm{i}V_{-1}\end{cases}$$

得：Y_1,Y_{-1}分别也是$(V^{+})_1$及$(V^{+})_{-1}$的标正基.

综上所述有：

$$V^{+}=(V^{+})_1\oplus(V^{+})_{-1}\oplus\overbrace{(V^{+})_{\lambda_1}\oplus(V^{+})_{\bar\lambda_1}}\oplus\cdots\oplus\overbrace{(V^{+})_{\lambda_l}\oplus(V^{+})_{\bar\lambda_l}}$$

取V_1的$\downarrow$标正基 …… 取标$\downarrow$正基 ……

$$Y_1\quad\cdots\cdots\quad\begin{cases}\eta_{\lambda_1 1}=\alpha_{\lambda_1 1}+\mathrm{i}\beta_{\lambda_1 1}\\ \eta_{\lambda_1 k_{\lambda_1}}=\alpha_{\lambda_1 k_{\lambda_1}}+\mathrm{i}\beta_{\lambda_1 k_{\lambda_1}}\end{cases}\quad\cdots\cdots$$

则有：

$$Y_1\cup Y_{-1}\cup\{\sqrt2\alpha_{\lambda_1 1},\cdots,\sqrt2\alpha_{\lambda_1 k_{\lambda_1}}\}\cup\{\sqrt2\beta_{\lambda_1 1},\cdots,\sqrt2\beta_{\lambda_1 k_{\lambda_1}}\}\cup\cdots$$
$$\cup\{\sqrt2\alpha_{\lambda_l 1},\cdots,\sqrt2\alpha_{\lambda_l k_{\lambda_l}}\}\cup\{\sqrt2\beta_{\lambda_l 1},\cdots,\sqrt2\beta_{\lambda_l k_{\lambda_l}}\}$$

构成V的一组标正基，$\mathbb{U}$在基：

$$Y_1\cup Y_{-1}\cup\{\sqrt2\alpha_{\lambda_1 1},\sqrt2\beta_{\lambda_1 1};\cdots;\sqrt2\alpha_{\lambda_1 k_{\lambda_1}},\sqrt2\beta_{\lambda_1 k_{\lambda_1}}\}\cup\cdots\cup\{\sqrt2\alpha_{\lambda_l 1},\sqrt2\beta_{\lambda_l 1};\cdots;\sqrt2\alpha_{\lambda_l k_{\lambda_l}},\sqrt2\beta_{\lambda_l k_{\lambda_l}}\}$$

下的表示阵为

$$\begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & \cos\theta & \sin\theta & \\ & & & & & & -\sin\theta & \cos\theta & \\ & & & & & & & & \ddots \end{pmatrix}$$

这就是正交矩阵在正交相似下的标准型.

(Ⅳ)秩为 1 的自伴变换的表示阵

对有限维内积空间$\langle V,\langle\cdot\rangle\rangle$上的一个线性变换$\mathbb{A}$,记其在 V 的一组标正基下的表示阵为 A.若满足$\mathbb{A}^* = \mathbb{A}$,则$\mathbb{A}$ 称为一个**自伴变换**.此时 A 为熟知的自共轭阵.一般地,我们知道有下列对应:

(1)严格正变换(即$\mathbb{A}$ 自伴且$\langle \mathbb{A}x,x\rangle > 0, \forall\, 0\neq x\in V$)$\longleftrightarrow A$ 为半正定阵;

(2)正变换(即$\mathbb{A}$ 自伴且$\langle \mathbb{A}x,x\rangle \geqslant 0, \forall\, 0\neq x\in V$)$\longleftrightarrow A$ 为半正定阵;

(3)规范变换(即$\mathbb{A}^*\mathbb{A} = \mathbb{A}\mathbb{A}^*$)$\longleftrightarrow A$ 为规范阵(即可酉相似对角化的矩阵).

定理 1.3.3　设$\mathbb{A}$ 为秩是 1 的 n 维内积空间$(V,\langle\cdot\rangle)$上的自伴变换,则$\mathbb{A}$ 在 V 的每个标正基下的表示阵 A 均可写成

$$k\cdot\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}(\bar{\beta}_1 \quad \cdots \quad \bar{\beta}_n)\ (0\neq k\in\mathbb{R},\beta_1,\cdots,\beta_n \text{ 不全为 } 0)$$

之形式.反之,给定 n 维内积空间$(V,\langle\cdot\rangle)$上的一个线性变换$\mathbb{A}$,若$\mathbb{A}$ 在 V 的某标正基下的表示阵具有上述之形式,则$\mathbb{A}$ 必为秩是 1 的自伴变换.

证明: 只需证前半部分.由于 A 为秩是 1 的自共轭阵,故 A 可酉相似于对角阵:

$$A = U\begin{pmatrix} k & 0\\ 0 & 0\end{pmatrix}U^*$$

其中 k 为实数.令 U 的第一列为$\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}$,则

$$A = kU\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}U^* = k\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}(\bar{\beta}_1 \quad \cdots \quad \bar{\beta}_n)$$

□

该定理 1.3.3 有一个直接的推论:

推论 1.3.1　(Hadamard 定理)两个半正定阵的 Hadamard 积仍为半正定阵.

证明: 设 A,B 均为半正定阵,秩 $A = r$,秩 $B = s$,则 A 可分解为 r 个秩为 1 的半正定

阵之和，B 可分解为 s 个秩为 1 的半正定阵之和.

因此，只须对两个秩为 1 的半正定阵证明其 Hadamard 积仍为半正定阵即可.故不妨设秩 $A=1$，秩 $B=1$.

由定理 1.3.3 知：

$$A=(a_{ij})=\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}(\bar{\beta}_1 \quad \cdots \quad \bar{\beta}_n)$$

$$B=(b_{ij})=\begin{pmatrix}\gamma_1\\ \vdots\\ \gamma_n\end{pmatrix}(\bar{\gamma}_1 \quad \cdots \quad \bar{\gamma}_n)$$

$$a_{ij}\cdot b_{ij}=\beta_i\bar{\beta}_j\cdot\gamma_i\bar{\gamma}_j=\beta_i\gamma_i\,\overline{\beta_j\gamma_j}$$

结果，$A\cdot B=(a_{ij}\cdot b_{ij})=\begin{pmatrix}\beta_1\gamma_1\\ \vdots\\ \beta_n\gamma_n\end{pmatrix}(\overline{\beta_1\gamma_1} \quad \cdots \quad \overline{\beta_n\gamma_n})$为半正定阵.

□

习题一

1. 任给单位向量 $\omega\in\mathbb{C}^{n\times1}$，$\|\omega\|=1$，矩阵 $H(\omega)=I_n-2\omega\omega^*$ 称为一个初等 Householder 阵或 Householder 变换阵.证明初等 Householder 阵具有下列性质：

(1)$H(\omega)$为自共轭酉阵，行列式为 -1；

(2)$\forall\alpha,\beta\in\mathbb{C}^{n\times1}(n>1)$，存在 n 阶 Householder 阵 $H(\omega)$使 $H(\omega)\alpha=\beta$ 的充要条件是等式组$\begin{cases}\alpha^*\alpha=\beta^*\beta\\ \alpha^*\beta=\beta^*\alpha\end{cases}$成立.

2. 设 $\alpha\in\mathbb{C}^n$.证明：存在 Householder 阵 $H(\omega)$使得 $H(\omega)\alpha=a\cdot e_1$，其中 $a\in\mathbb{C}$，$e_1=(1,0,\cdots,0)^{\mathrm{T}}\in\mathbb{C}^{n\times1}$.

3. 设 A,B,C 是 n 阶复方阵，满足 $CAA^*=BAA^*$，证明 $CA=BA$.

4. 设 $A\in\mathbb{C}^{n\times n}$，证明 Hadamard 不等式：$|\det A^2|\leqslant\prod_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2$.

5. 设 U 为酉阵，$A=\begin{pmatrix}a_1 & & \\ & \ddots & \\ & & a_n\end{pmatrix}$.证明：$UA$ 的特征值μ 满足 $m\leqslant|\mu|\leqslant M$，其中 $m=\min\limits_{1\leqslant i\leqslant n}|a_i|$，$M=\max\limits_{1\leqslant i\leqslant n}|a_i|$.

6. 证明：复数域上每个正规矩阵 $A=(a_{ij})$至少有一个特征值 λ_t 满足 $|\lambda_t|\geqslant\max\limits_{i,j}|a_{ij}|$.

7. 证明：每个交错的 k－线性型必为斜对称的；反之，当线性空间 V 的系数域F 的特

征不为 2 时,每个斜对称的 k - 线性型也必为交错 k - 线性型.

8. 设 B 是正定矩阵,A 为与 B 同阶的矩阵.若有 $\alpha\neq 0$ 满足 $A\alpha=\lambda B\alpha$,则称 λ 是 A 关于 B 的相应特征值,α 是使 A 关于 B 的相应于 λ 的相对特征向量.试证明:当 A 是 Hermite 矩阵时,A 关于 B 的相应特征值为实数且相应的相对特征向量可组成完全集.

9. 设 B 是实正定矩阵,S 是实反对称矩阵.证明 $\det(B+S)\geqslant \det B$.

10. 设 G 是 $\mathbb{R}$ 上 n 阶非奇异矩阵全体,H 是由 n 阶对称矩阵组成的 G 的乘法子群.证明存在 H 中所有矩阵共有的非零特征向量.

11. 设 B 是实正交正定矩阵,则 $B=I$.

12. 设 n 个不全为 0 的实常数为 $a_1,a_2,\cdots,a_n$.求 n^2 个未知数 $x_{jk}(j,k=1,2,\cdots,n)$ 的如下线性方程组的解空间:$x_{jk}a_ia_l-x_{il}a_ja_k=0(i,j,l,k=1,2,\cdots,n)$.

13. 证明:酉空间中两个可换的酉算子有公共特征向量组成的标正基.

第2章　矩阵分解

给定 $m\times n$ 阶矩阵A，A 可化为$A=P\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}Q$，其中 $r=\mathrm{rank}A$，P，Q 皆可逆．显然 A 还可以写成 $A=P\begin{pmatrix} I_r\\ 0\end{pmatrix}(I_r\quad 0)Q$．让 P 的前r列记为G，Q 的前r行记为H，于是 $A=GH$，GH 便是我们熟知的满秩分解，它是众多矩阵分解中的一种．综观矩阵代数，矩阵分解是解剖矩阵结构的重要手段之一，其本质是，通过建立相应的矩阵分解，从而使某些问题得以简化和分解，进而更加清晰地获取矩阵的相关特性．本章主要讲述几种常见而又重要的矩阵分解，他们中有些在矩阵计算中发挥重要的作用，有些则在理论证明中扮演着关键的角色．

2.1　Jordan 分解与 Frobenius 分解

矩阵的 Jordan 分解定理在某种意义上可看成是矩阵分解理论的制高点，其定理内容如下：

定理 2.1.1　设 $A\in\mathbb{C}^{n\times n}$，则存在可逆矩阵 $T\in\mathbb{C}^{n\times n}$使得

$$A=TJT^{-1}$$

其中 $J=\begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_k}(\lambda_k)\end{pmatrix}$，$J_{n_i}(\lambda_i)=\begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{n_i\times n_i}$，$n_1+\cdots+n_k=n$，

$\lambda_1,\cdots,\lambda_k$ 可以相同，也可以不同．进一步，J 除了若当块顺序外是唯一的．

一般地，称 $A=TJT^{-1}$为 A 的 **Jordan 分解**，J 是A 的 **Jordan 标准型**．该定理的证明可以有多种途径：

途径一：使用 λ－阵理论．

设 $A,B\in\mathbb{C}^{n\times n}$，则在 $\mathbb{C}$ 上有：

$A\sim B\Leftrightarrow\lambda I-A$ 与 $\lambda I-B$ 在$\mathbb{C}$ 上等价；

$\Leftrightarrow A$ 与 B 有相同的不变因子组；

$\Leftrightarrow\lambda I-A$ 与 $\lambda I-B$ 有相同的行列式因子组；

$\Leftrightarrow$ 在$\mathbb{C}$ 上 A 与 B 有相同的初等因子组．

设 A 的初等因子组为：$(\lambda-\lambda_1)^{n_1},(\lambda-\lambda_2)^{n_2},\cdots,(\lambda-\lambda_s)^{n_s}$，其中 $\lambda_1,\cdots,\lambda_s$ 可以相同，也可以不同；同样，$n_1,\cdots,n_s$ 可以相同，也可以不同．那么每一个初等因子 $(\lambda-\lambda_i)^{n_i}$ 对应于一个 Jordan 块：

$$J_{n_i}(\lambda_i)=\begin{pmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i\end{pmatrix}$$

于是，这些 Jordan 块构成一个 Jordan 阵：

$$J=\begin{pmatrix}J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_s}(\lambda_s)\end{pmatrix}$$

直接计算可知，J 的初等因子组与 A 的初等因子组是相同的，故 J 与 A 相似．倘若另一个 Jordan 阵 J' 与 A 相似，则由于 J' 与 A 有相同的初等因子组，从而 J' 与 J 除了其中 Jordan块顺序外是相同的．

途径二：使用线性变换语言．

首先，令 A 为 $V=\mathbb{C}^n$ 在选定标正基 $\{e_1,\cdots,e_n\}$ 下所代表的线性变换 α，取 A 的特征多项式的标准分解为 $f(\lambda)=(\lambda-\lambda_1)^{r_1}\cdots(\lambda-\lambda_s)^{r_s}$，则 V 可分解为 α 的不变子空间的直和：

$$V=V_1\oplus\cdots\oplus V_s$$

其中 $V_i=\{\zeta\mid(\alpha-\lambda_i I)^{r_i}\zeta=0,\zeta\in V\}$，$i=1,2,\cdots,s$．现让 $\beta_i=(\alpha-\lambda_i I)|_{V_i}$，则 β_i 幂零，$\beta_i^{r_i}=0$．由 β_i 幂零知，V_i 中存在如下形式的一组基（其推导可参见北大《高等代数》）：

$$\begin{array}{cccc}\alpha_1 & \alpha_2 & \cdots & \alpha_t \\ \beta_i\alpha_1 & \beta_i\alpha_2 & \cdots & \beta_i\alpha_t \\ \vdots & \vdots & & \vdots \\ \beta_i^{k_1-1}\alpha_1 & \beta_i^{k_2-1}\alpha_2 & \cdots & \beta_i^{k_t-1}\alpha_t\end{array}$$

$$(\beta_i^{k_1}\alpha_1=0,\beta_i^{k_2}\alpha_2=0,\cdots,\beta_i^{k_t}\alpha_t=0)$$

于是 β_i 在该组基下的表示阵可为

$$\begin{pmatrix}\underbrace{\begin{matrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0\end{matrix}}_{k_1} & & \\ & \ddots & \\ & & \underbrace{\begin{matrix}0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0\end{matrix}}_{k_t}\end{pmatrix}$$

从而$\alpha|_{V_i}=\beta_i+\lambda_i I$在该基下的表示阵为：

$$\begin{pmatrix} \underbrace{\begin{matrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i\end{matrix}}_{k_1} & & \\ & \ddots & \\ & & \underbrace{\begin{matrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i\end{matrix}}_{k_t} \end{pmatrix}$$

把每个 V_i 上的上述基合起来构成V的基，则 α 在该基下的表示阵为一个 Jordan 标准型.

其次，设λ为A的特征值，证明A的 Jordan 标准型 J 中以λ为特征值、阶为 l 的Jordan块个数是

$$r_{l+1}+r_{l-1}-2r_l$$

其中 $r_l=\text{rank}(\lambda I-A)^l$.由于 $r_{l+1}+r_{l-1}-2r_l$ 只跟A与λ有关，从而A的 Jordan 标准型的唯一性获证.

事实上，因为 $A=TJT^{-1}$，则有 $\lambda I-A=T(\lambda I-J)T^{-1}$.对于任何正整数 l，均有$(\lambda I-A)^l=T(\lambda I-J)^lT^{-1}$，从而 $\text{rank}(\lambda I-A)^l=\text{rank}(\lambda I-J)^l$.但

$$(\lambda I-J)^l=\begin{pmatrix}(\lambda I-J_{n_1}(\lambda_1))^l & & \\ & \ddots & \\ & & (\lambda I-J_{n_s}(\lambda_S))^l\end{pmatrix}$$

因此，对$(\lambda I-J)^l$中的每一个对角块$(\lambda I-J_{n_i}(\lambda_i))^l$均有：

$$\begin{cases}\text{若 } \lambda\neq\lambda_i\text{，则 } \text{rank}(\lambda I-J_{n_i}(\lambda_i))^l=n_i \\ \text{若 } \lambda=\lambda_i\text{，则 } \text{rank}(\lambda I-J_{n_i}(\lambda_i))^l=\begin{cases}n_i-l, & \text{当 } l<n_i \text{ 时} \\ 0, & \text{当 } l\geqslant n_i \text{ 时}\end{cases}\end{cases}$$

结果有

$$\text{rank}(\lambda I-J_{n_i}(\lambda_i))^l-\text{rank}(\lambda I-J_{n_i}(\lambda_i))^{l+1}=\begin{cases}0,\lambda_i\neq\lambda \\ 0,\lambda_i=\lambda \text{ 且 } l\geqslant n_i \\ 1,\lambda_i=\lambda \text{ 且 } l<n_i\end{cases}$$

考察 J 的对角块元：$J_{n_1}(\lambda_1),\cdots,J_{n_s}(\lambda_s)$.这些对角块元中以$\lambda$为特征值，阶数大于 l 的每个 $J_{n_i}(\lambda_i=\lambda)$，均有

$$\text{rank}(\lambda I-J_{n_i}(\lambda_i))^l-\text{rank}(\lambda I-J_{n_i}(\lambda_i))^{l+1}=1$$

因此

$$\begin{aligned}r_l - r_{l+1} &= \operatorname{rank}(\lambda I - J)^l - \operatorname{rank}(\lambda I - J)^{l+1} \\ &= \sum_{i=1}^{s}(\operatorname{rank}(\lambda I - J_{n_i}(\lambda_i))^l - \operatorname{rank}(\lambda I - J_{n_i}(\lambda_i))^{l+1}) \\ &= J \text{ 中,以 } \lambda \text{ 为特征值且阶数大于 } l \text{ 的 Jordan 块的个数}\end{aligned}$$

同理

$$r_{l-1} - r_l = J \text{ 中,以 } \lambda \text{ 为特征值且阶数大于 } l-1 \text{ 的 Jordan 块的个数}$$

故 J 中以 λ 为特征值且阶数等于 l 的 Jordan 块的个数是

$$r_{l-1} - r_l - (r_l - r_{l+1}) = r_{l+1} + r_{l-1} - 2r_l$$

途径三:借助于空间分解理论,其本质实际上与途径二相同.

设 $A \in \mathbb{C}^{n\times n}$,$\lambda$ 为 A 的一个特征值,于是升链

$$N(A-\lambda I) \subseteq \cdots \subseteq N(A-\lambda I)^k \subseteq \cdots$$

必有限.

设首次出现 $N(A-\lambda I)^k = N(A-\lambda I)^{k+1}$ 的 k 为 r,此时,对 $N(A-\lambda I)^r$ 为 A 的相应于 λ 的**广义特征子空间**,r 称为 λ 的**指标**或 $A-\lambda I$ 的**指标**.对每个自然数 $k \leqslant r$,称 $N(A-\lambda I)^k \setminus N(A-\lambda I)^{k-1}$ 中的元为 A 的相应于 λ 的秩为 k 的**广义特征向量**.

易见:

(1)A 的相应于 λ 的秩为 1 的广义特征向量即为 A 的相应于 λ 的特征向量.

(2)A 中秩各不相同的相应于 λ 的广义特征向量必线性无关.

(3)记 $N_\lambda = N(A-\lambda I)^r$,$\mathrm{Im}_\lambda = \mathrm{Im}(A-\lambda I)^r$,其中 r 为 λ 的指标,则有 $\mathbb{C}^n = N_\lambda \oplus \mathrm{Im}_\lambda$,且 N_λ,Im_λ 均为 A 的不变子空间.

(4)若 λ_1,λ_2 分别为 A 的两个互异特征值,它们的指标值分别为 r_{λ_1},r_{λ_2},则有 $N_{\lambda_2} \subseteq \mathrm{Im}_{\lambda_1}$,$N_{\lambda_1} \subseteq \mathrm{Im}_{\lambda_2}$.

(5)设 $\lambda_1,\cdots,\lambda_s$ 为 A 的全部互异特征值,则有:$\mathbb{C}^n = N_{\lambda_1} \oplus N_{\lambda_2} \oplus \cdots \oplus N_{\lambda_s}$.现在 N_{λ_1},N_{λ_2},$\cdots$,N_{λ_s} 中任取一元,记为 N_λ,我们进一步对 N_λ 施行分解.

记 r 为 λ 的指标,让 $N^j = N(A-\lambda I)^j$,$j=1,2,\cdots,r$,则有 $N^1 \subset N^2 \subset \cdots \subset N^r = N_\lambda$,于是选取 N^1 的一组基 β^1,扩充为 N^2 的基 $\beta^1 \cup \beta^2$,则 β^2 是由秩为 2 的一些广义特征向量构成的无关组;将 $\beta^1 \cup \beta^2$ 扩充为 N^3 的基 $\beta^1 \cup \beta^2 \cup \beta^3 \cdots$ 直到扩充为 N_λ 的基:$\beta^1 \cup \beta^2 \cup \beta^3 \cdots \cup \beta^r$,于是 $\beta^i(i=1,2,\cdots,r)$ 是由秩 i 的一些广义特征向量构成的无关组,结果 N_λ 被分解为 $N_\lambda = \mathrm{span}\beta^1 \oplus \mathrm{span}\beta^2 \oplus \cdots \oplus \mathrm{span}\beta^r$,其中 $\mathrm{span}\beta^1 = N^1$.显然 $\mathrm{span}\beta^1$,$\mathrm{span}\beta^2$,$\cdots$,$\mathrm{span}\beta^r$ 的维数分别为:

$$\begin{gathered}d_1 = \dim\mathrm{span}\beta^1 = \dim N^1 - 0 \\ d_2 = \dim\mathrm{span}\beta^2 = \dim N^2 - \dim N^1 \\ \vdots \\ d_r = \dim\mathrm{span}\beta^r = \dim N_\lambda - \dim N^{r-1}\end{gathered}$$

(6)选取 $\mathrm{span}\beta^{r}$ 的基 $\alpha_1,\cdots,\alpha_{d_r}$ 后,必有 $\mathrm{span}\beta^{r-1}$ 的一组基可选为:

$$(A-\lambda I)\alpha_1,\cdots,(A-\lambda I)\alpha_{d_r};\alpha_{d_r+1},\cdots,\alpha_{d_{r-1}}$$

继而 $\mathrm{span}\beta^{r-2}$ 的一组基可选为:

$$(A-\lambda I)^2\alpha_1,\cdots,(A-\lambda I)^2\alpha_{d_r};(A-\lambda I)\alpha_{d_r+1},\cdots,(A-\lambda I)\alpha_{d_{r-1}};\alpha_{d_{r-1}+1},\cdots,\alpha_{d_{r-2}}$$

最后 $\mathrm{span}\beta^{1}=N^{1}$ 的一组基可选为:

$$(A-\lambda I)^{r-1}\alpha_1,\cdots,(A-\lambda I)^{r-1}\alpha_{d_r};(A-\lambda I)^{r-2}\alpha_{d_r+1},\cdots,(A-\lambda I)^{r-2}\alpha_{d_{r-1}};\cdots;\alpha_{d_2+1},\cdots,\alpha_{d_1}$$

(7)将上述中的向量重排如下:

$$\begin{cases}\alpha_1,(A-\lambda I)\alpha_1,(A-\lambda I)^2\alpha_1,\cdots,(A-\lambda I)^{r-1}\alpha_1\\ \vdots\\ \alpha_{d_r},(A-\lambda I)\alpha_{d_r},(A-\lambda I)^2\alpha_{d_r},\cdots,(A-\lambda I)^{r-1}\alpha_{d_r}\end{cases}$$

$$\begin{cases}\alpha_{d_r+1},(A-\lambda I)\alpha_{d_r+1},(A-\lambda I)^2\alpha_{d_r+1},\cdots,(A-\lambda I)^{r-2}\alpha_{d_r+1}\\ \vdots\\ \alpha_{d_r-1},(A-\lambda I)\alpha_{d_r-1},(A-\lambda I)^2\alpha_{d_r-1},\cdots,(A-\lambda I)^{r-2}\alpha_{d_r-1}\end{cases}$$

$$\vdots$$

$$\begin{cases}\alpha_{d_3+1},(A-\lambda I)\alpha_{d_3+1}\\ \vdots\\ \alpha_{d_2},(A-\lambda I)\alpha_{d_2}\end{cases}$$

$$\{\alpha_{d_2+1},\cdots,\alpha_{d_1}$$

则若令

$$P=(\alpha_1,(A-\lambda I)\alpha_1,(A-\lambda I)^2\alpha_1,\cdots,(A-\lambda I)^{r-1}\alpha_1;\cdots;\\ \alpha_{d_r},(A-\lambda I)\alpha_{d_r},(A-\lambda I)^2\alpha_{d_r},\cdots,(A-\lambda I)^{r-1}\alpha_{d_r};\cdots;\alpha_{d_2+1},\cdots,\alpha_{d_1})$$

有 P 为 $n\times\dim N_\lambda$ 的列满秩阵,且

$$AP=P\begin{pmatrix}\lambda &&&&&&&&&&\\ 1&\ddots&&&&&&&&&\\ &\ddots&\ddots&&&&&&&&\\ &&1&\lambda&&&&&&&\\ &&&&\ddots&&&&&&\\ &&&&&\lambda&&&&&\\ &&&&&1&\ddots&&&&\\ &&&&&&\ddots&\ddots&&&\\ &&&&&&&1&\lambda&&\\ &&&&&&&&&\lambda&\\ &&&&&&&&&&\ddots\\ &&&&&&&&&&&\lambda\end{pmatrix}$$

现分别取 N_λ 为$\mathbb{C}^n = N_{\lambda_1} \oplus \cdots \oplus N_{\lambda_s}$ 中的 $N_{\lambda_1}, \cdots, N_{\lambda_s}$，按上述途径获取的 $P_1, P_2, \cdots, P_s$，则有

$$A(P_1, P_2, \cdots, P_s) = (P_1, P_2, \cdots, P_s) \cdot J$$

其中$(P_1, P_2, \cdots, P_s)$为 n 阶可逆方阵，A 相似于 Jordan 标准型 J.

由上述过程知，相应于 A 的特征值λ，其 1 阶 Jordan 块个数有 $d_1 - d_2$ 个，2 阶 Jordan 块个数有 $d_2 - d_3$ 个，……，$r-1$ 阶 Jordan 块个数有 $d_{r-1} - d_r$ 个；r 阶 Jordan 块个数有 d_r 个. 因此，理论上途径三不仅给出了方阵 A 的 Jordan 标准型的求法，同时还给出了变换矩阵 P 的求法，步骤简述如下：

第一步：求出 A 的所有不同特征值$\lambda_1, \cdots, \lambda_s$.

第二步：确定每个特征值 λ 的指标 r，进而求出 $d_1 - d_2, d_2 - d_3, \cdots, d_{r-1} - d_r, d_r$，从而决定出该 λ 的 1 阶，2 阶，……，r 阶的 Jordan 块个数.

第三步：将第二步所得的 Jordan 块排在主对角线上，即得 A 的 Jordan 标准型 J.

第四步：像前述(6)中一样选出 N_λ 的基(列向量集)，并将所求出的各 N_λ 的基按 J 中对应子块排定次序组成 P，P 即为所求的变换阵.

关于矩阵 A 的 Jordan 标准型的求法，熟知的途径是：利用 A，通过初等变换，求出 $\lambda I - A$的 Smith 标准型$\begin{pmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_s(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$，进而求出 A 的初等因子组，直接得到其 Jordan 标准型. 这里指出，上述中的途径二，有时对我们求取 J 能带来较大方便.

例 2.1.1　设 $A = \begin{pmatrix} 4 & -1 & -1 & 0 \\ 4 & 0 & -2 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 6 & 1 \end{pmatrix}$，求 A 的 Jordan 标准型.

解： 先求 A 的特征值，$\det(\lambda I - A) = (\lambda - 2)^3(\lambda - 1)$，故 $\lambda_1 = 2$ 为 3 重根，$\lambda_2 = 1$ 为 1 重根.

对 $\lambda_2 = 1$，其代数重数为 1，从而几何重数为 1，故 A 相应于特征值 1 的 Jordan 块只有 1 个，它为 1 阶.

对 $\lambda_1 = 2$，其 1 阶 Jordan 块个数为：

$$\text{rank}(2I_4 - A)^2 + \text{rank}(2I_4 - A)^0 - 2\text{rank}(2I_4 - A) = 1$$

由于 A 为 4 阶方阵，因此相应于特征值 2 的 2 阶 Jordan 块个数必为 1，所以 A 的 Jordan 标

准型为$\begin{pmatrix} 1 & & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{pmatrix}$.

□

矩阵分解中常被人们关心的问题是矩阵分解的封闭性问题.我们知道矩阵的 Jordan 分解对实数域不是封闭的.下面来研究实矩阵的 Jordan 分解形式.

定义 2.1.1 形如

$$C_k(a,b)=\begin{pmatrix} C(a,b) & I_2 & & \\ & C(a,b) & \ddots & \\ & & \ddots & I_2 \\ & & & C(a,b) \end{pmatrix}$$

其中 $C(a,b)=\begin{pmatrix} a & b \\ -b & a \end{pmatrix}(a,b\in\mathbb{R})$的 $2k\times 2k$ 矩阵叫作一个**实 Jordan 块**.由若干实的 Jordan 块构成的 n 阶对角阵:

$$J_{\mathbb{R}}=\begin{pmatrix} C_{n_1}(a_1,b_1) & & & & & \\ & \ddots & & & & \\ & & C_{n_p}(a_p,b_p) & & & \\ & & & J_{n_q}(\lambda_q) & & \\ & & & & \ddots & \\ & & & & & J_{n_r}(\lambda_r) \end{pmatrix}$$

其中 $C_{n_i}(a_i,b_i)(i=1,2,\cdots,p)$为具有共轭特征值(非实数)的实 Jordan 块,$J_{n_j}(\lambda_j)(j=q,q+1,\cdots,r)$为对角元是实数的 Jordan 块,称 J_R 为**实 Jordan 阵**.

定理 2.1.2 设 $A\in\mathbb{R}^{n\times n}$,则存在可逆 $T\in\mathbb{R}^{n\times n}$使得 $A=TJ_RT^{-1}$,其中 J_R 是一个实 Jordan 阵.

通常称此定理中的 $A=TJ_RT^{-1}$为 A 的**实 Jordan 分解**,其中的 J_R 称为 A 的**实 Jordan 标准型**.

证明: 若 $a+b\mathrm{i}$ 是 A 的一个复特征值($b\neq 0$),$J_k(a+b\mathrm{i})$为 A 的一个 Jordan 块,向量 $x_1+y_1\mathrm{i},\cdots,x_k+y_k\mathrm{i}$(其中 $x_l,y_l\in\mathbb{R}^n$)满足

$$A(x_1+y_1\mathrm{i},\cdots,x_k+y_k\mathrm{i})=(x_1+y_1\mathrm{i},\cdots,x_k+y_k\mathrm{i})J_k(a+b\mathrm{i})$$

则易知有

$$A(x_1-y_1\mathrm{i},\cdots,x_k-y_k\mathrm{i})=(x_1-y_1\mathrm{i},\cdots,x_k-y_k\mathrm{i})J_k(a-b\mathrm{i})$$

因此有:

$$A(x_1,y_1)=(x_1,y_1)\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

$$A(x_2,y_2)=(x_2,y_2)\begin{pmatrix} a & b \\ -b & a \end{pmatrix}+(x_1,y_1)$$

$$\vdots$$

$$A(x_k,y_k)=(x_k,y_k)\begin{pmatrix} a & b \\ -b & a \end{pmatrix}+(x_{k-1},y_{k-1})$$

结果,由$\{x_1+y_1\mathrm{i},\cdots,x_k+y_k\mathrm{i},x_1-y_1\mathrm{i},\cdots,x_k-y_k\mathrm{i}\}$为线性无关组知$\{x_1,y_1,x_2,y_2,\cdots,x_k,y_k\}$也为实线性无关组,进一步,矩阵$(x_1,y_1;x_2,y_2;\cdots;x_k,y_k)$恰与 $C_k(a,b)$相对应,定理获证.

□

由于

$$C_k(a,b)\sim\begin{pmatrix} J_k(a+b\mathrm{i}) & \\ & J_k(a-b\mathrm{i}) \end{pmatrix}$$

其中

$$J_k(a+b\mathrm{i})=\begin{pmatrix} a+b\mathrm{i} & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & a+b\mathrm{i} \end{pmatrix}_{k\times k}$$

因此一个 $2k$ 阶具有共轭特征值的实 Jordan 块与两个 k 阶特征值共轭的 Jordan 块构成的对角块阵是相似的,从而矩阵的实 Jordan 标准型在本质上也是唯一的.

定理 2.1.1 说明的是复数域上的任一矩阵 A 必相似于一个 Jordan 型矩阵.设$\mathbb{F}$ 为任意数域,下面证明$\mathbb{F}$ 上任一矩阵必相似于$\mathbb{F}$ 上的一个**有理标准型矩阵**.有理标准型矩阵又称 **Frobenius 阵**.

定义 2.1.2　对数域$\mathbb{F}$ 上的一个多项式 $f(\lambda)=\lambda^n+a_1\lambda^{n-1}+\cdots+a_n$,称矩阵

$$A=\begin{pmatrix} 0 & & & -a_n \\ 1 & \ddots & & \vdots \\ & \ddots & \ddots & -a_2 \\ & & 1 & -a_1 \end{pmatrix}$$

为多项式 $f(\lambda)$的**友矩阵**.

例 2.1.2　设 A 为定义 2.1.2 中的矩阵,求它的特征矩阵 $\lambda I-A$ 的不变因子,并将它化为 Smith 标准型.

解: $\lambda I-A=\begin{pmatrix} \lambda & & & a_n \\ -1 & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ & & -1 & \lambda+a_1 \end{pmatrix}$,将 $\lambda I-A$ 的第 2,3,$\cdots$,n 行分别乘以 $\lambda,\lambda^2,\cdots,\lambda^{n-1}$后都加到第 1 行上去,得

$$\lambda I - A \sim \begin{pmatrix} 0 & & & & f(\lambda) \\ -1 & \lambda & & & a_{n-1} \\ & \ddots & \ddots & & \vdots \\ & & \ddots & \lambda & a_2 \\ & & & -1 & \lambda + a_1 \end{pmatrix}$$

于是

$\det(\lambda I - A) = f(\lambda)$,从而 $D_n(\lambda) = f(\lambda), D_{n-1}(\lambda) = 1, \cdots, D_0(\lambda) = 1$.

结果

$$d_n(\lambda) = \frac{D_n(\lambda)}{D_{n-1}(\lambda)} = f(\lambda), d_{n-1}(\lambda) = \cdots = d_1(\lambda) = 1$$

故 A 的 Smith 标准型为$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & f(\lambda) \end{pmatrix}$,不变因子为 $1,1,\cdots,1,f(\lambda)$.

□

定义 2.1.3 下列准对角阵 $A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_s \end{pmatrix}$,其中 A_i 分别是数域 $\mathbb{F}$ 上某些首 1 多项式 $d_i(\lambda)(i = 1,2,\cdots,s)$的友矩阵且满足 $d_1(\lambda) \mid d_2(\lambda) \mid \cdots \mid d_s(\lambda)$,则称 A 为$\mathbb{F}$ 上的 **Frobenius 标准型矩阵(或有理标准型矩阵)**.

定理 2.1.3 数域$\mathbb{F}$ 上任何 $n \times n$ 方阵A 在 $\mathbb{F}$ 上均相似于唯一的一个有理标准型,称为 A 的 **Frobenius 标准型(或有理标准型)**.

为证此定理,先证下列引理,该引理剖析了有理标准型矩阵的不变因子性质.

引理 2.1.1 设 $B = \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_s \end{pmatrix}$为$\mathbb{F}$ 上的一个有理标准型,则 B 的不变因子为 $1,1,\cdots,1,d_1(\lambda),d_2(\lambda),\cdots,d_s(\lambda)$,其中 1 的个数等于 $\deg d_1(\lambda) + \cdots + \deg d_s(\lambda) - s$.

证明: 由 B_i 的不变因子组为 $1,1,\cdots,1,d_i(\lambda)$直接可知.

□

定理 2.1.3 的证明: 设 A 的不变因子为

$$1,1,\cdots,1,d_1(\lambda),d_2(\lambda),\cdots,d_s(\lambda),$$

其中 $d_1(\lambda) \mid d_2(\lambda) \mid \cdots \mid d_s(\lambda)$, $d_1(\lambda),d_2(\lambda),\cdots,d_s(\lambda)$的次数均$\geq 1$.显见 1 的个数为 $n-s$,记 $d_i(\lambda)$的友矩阵为 B_i,考察矩阵

$$B = \begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_s \end{pmatrix}$$

于是 B 的不变因子与 A 的不变因子相同,从而 A 与 B 相似.注意到 A 的不变因子组是由

A 唯一确定的,故 A 在 $\mathbb{F}$ 上只有相似于唯一的一个有理标准型.

□

通常,数域$\mathbb{F}$上的一个矩阵 A 被称作是**非减次矩阵**是指A的任一特征值的几何重数为1.否则称矩阵 A 为**减次的**.

定理 2.1.4　设 $A\in\mathbb{F}^{n\times n}$, C 为 A 的特征多项式的友矩阵,则以下两条等价:

(1) C 与 A 相似;

(2) A 是非减次矩阵.

证明: (1)⇒(2)　由 C 的不变因子组为 $1,1,\cdots,1,\det(\lambda I-A)$ 立即知.

(2)⇒(1)　设 A 的 Jordan 标准型为

$$J=\begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_s}(\lambda_s)\end{pmatrix}$$

其中 $n_1+\cdots+n_s=n$, $\lambda_1,\cdots,\lambda_s$ 各不相同.又设 A 的特征多项式 $\det(\lambda I-A)$ 为 $\lambda^n+c_1\lambda^{n-1}+\cdots+c_n$,于是

$$C=\begin{pmatrix} 0 & & & -c_n \\ 1 & \ddots & & \vdots \\ & \ddots & \ddots & -c_2 \\ & & 1 & -c_1\end{pmatrix}$$

注意到 J 与 C 有相同的初等因子组,故 J 与 C 相似.

□

无论$\mathbb{F}=\mathbb{R}$还是$\mathbb{F}=\mathbb{C}$,由于(实)Jordan 型阵中每个(实)Jordan 块均是非减次矩阵,因此它相似于其对应的友矩阵,这实际上也再一次证明了$\mathbb{F}$上矩阵必相似于$\mathbb{F}$上的一个 Frobenius 标准型.进一步,利用 λ – 阵理论的结果,还可以使得变换矩阵仍属于$\mathbb{F}^{n\times n}$.

矩阵的 Frobenius 分解在理论上有着重要的作用,下面的几个结果初步说明了这一点.

推论 2.1.1　数域$\mathbb{F}$上矩阵 A 的 Frobenius 标准型为$\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_s\end{pmatrix}$, $\varphi(\lambda)$为$\mathbb{F}$上的多项式,则有 $\varphi(A)=0$ 当且仅当 $\varphi(A_s)=0$.

证明: 立知.

□

推论 2.1.2　设$\mathbb{F}$为数域, $A\in\mathbb{F}^{n\times n}$,则存在对称阵 $S,T\in\mathbb{F}^{n\times n}$使得 $A=ST$,其中 S, T 中至少有一个为可逆阵.

证明: 首先注意到存在可逆 $P\in\mathbb{F}^{n\times n}$使得 $A=PB_FP^{-1}$,其中 $B_F=\begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_s\end{pmatrix}$为

$\mathbb{F}$ 上的 Frobenius 标准型矩阵.现在分别对 $B_1, B_2, \cdots, B_s$ 进行分析.设 B_t 的阶数为 r,记之为

$$B_t = \begin{pmatrix} 0 & & & -b_r \\ 1 & \ddots & & \vdots \\ & \ddots & \ddots & -b_2 \\ & & 1 & -b_1 \end{pmatrix}$$

让

$$W_t = \begin{pmatrix} b_{r-1} & b_{r-2} & b_{r-3} & \cdots & b_1 & 1 \\ b_{r-2} & ⋰ & ⋰ & & ⋰ & \\ b_{r-3} & ⋰ & & & & \\ \vdots & & & & & \\ b_1 & & & & & \\ 1 & & & & & \end{pmatrix}$$

则有 W_t 为对称可逆阵,且通过直接计算可知 B_tW_t 为对称阵.于是令

$$W = \begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}$$

则

$$A = P\begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_s \end{pmatrix}P^{-1} = P\begin{pmatrix} B_1 & & \\ & \ddots & \\ & & B_s \end{pmatrix}\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\begin{pmatrix} W_1^{-1} & & \\ & \ddots & \\ & & W_s^{-1} \end{pmatrix}P^{-1}$$

$$= P\begin{pmatrix} B_1W_1 & & \\ & \ddots & \\ & & B_sW_s \end{pmatrix}P^{\mathrm{T}}\cdot(P^{-1})^{\mathrm{T}}\begin{pmatrix} W_1^{-1} & & \\ & \ddots & \\ & & W_s^{-1} \end{pmatrix}P^{-1}$$

再令

$$S = P\begin{pmatrix} B_1W_1 & & \\ & \ddots & \\ & & B_sW_s \end{pmatrix}P^{\mathrm{T}}$$

$$T = (P^{-1})^{\mathrm{T}}\begin{pmatrix} W_1^{-1} & & \\ & \ddots & \\ & & W_s^{-1} \end{pmatrix}P^{-1}$$

则有 S, T 皆为对称阵,$A = ST$,T 可逆.

当然,若对 A 采取如下分解:

$$A = P\begin{pmatrix} B_1W_1 & & \\ & \ddots & \\ & & B_sW_s \end{pmatrix}\begin{pmatrix} W_1^{-1} & & \\ & \ddots & \\ & & W_s^{-1} \end{pmatrix}P^{-1}$$

$$= P\left[P\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\right]^{\mathrm{T}}\left[P\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\right]^{-\mathrm{T}}$$

$$\cdot\begin{pmatrix} B_1W_1 & & \\ & \ddots & \\ & & B_sW_s \end{pmatrix}\left[P\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\right]^{-1}$$

再令

$$S = P\left[P\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\right]^{\mathrm{T}}$$

$$T = \left[P\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\right]^{-\mathrm{T}}\begin{pmatrix} B_1W_1 & & \\ & \ddots & \\ & & B_sW_s \end{pmatrix}\left[P\begin{pmatrix} W_1 & & \\ & \ddots & \\ & & W_s \end{pmatrix}\right]^{-1}$$

则有 S,T 皆为对称阵，$A = ST$，此时 S 可逆．

□

推论 2.1.3　设 $A\in\mathbb{F}^{n\times n}$，$\mathbb{F}$ 为任意数域，则存在可逆的对称阵 $T\in\mathbb{F}^{n\times n}$，使得 $A = TA^{\mathrm{T}}T^{-1}$．

2.2　矩阵的 Schur 分解与谱分解

矩阵的 Schur 分解是一种理论上存在，而实际上通常不方便通过有限次运算而获得，只能用迭代方法进行逼近的矩阵分解．其内容是：

定理 2.2.1　设 $A\in\mathbb{C}^{n\times n}$，则存在酉阵 U 使得 $UAU^* = R$，其中 R 为$\mathbb{C}$ 上的上三角阵．

由于任何一个实向量空间 V 上的线性变换必有维数不超过 2 的不变子空间，因此，利用归纳法立即可得实矩阵的 Schur 分解定理．

定理 2.2.2　设 $A\in\mathbb{R}^{n\times n}$，则存在正交阵 Q，使得 $QAQ^{\mathrm{T}} = R$，其中 R 形如：

$$\begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & \ddots & & \vdots \\ & & \ddots & \vdots \\ & & & R_{mm} \end{pmatrix}$$

为实分块阵，且 $R_{ii}(i = 1,2,\cdots,m)$或是 1 阶的或是有共轭复特征值的 2 阶实矩阵．

矩阵的 Schur 分解常在理论证明中发挥重要的作用．例如，以下几个结论均可由 Schur 分解定理直接推得．

结论 2.2.1　设 $A\in\mathbb{F}^{n\times n}$，则

A 为 Hermite 阵$\Leftrightarrow A = UDU^*$，其中 U 为酉阵，D 为实对角阵；

$\Leftrightarrow$ A 的特征值皆为实数且$\mathbb{F}^n$ 有以 A 的特征向量构成的一组标正基；

$\Leftrightarrow$ 存在可逆阵 V，使得 $A = VD_0V^*$，其中 $D_0 = \begin{pmatrix} I_k & & \\ & -I_{r-k} & \\ & & 0 \end{pmatrix}$，$k$ 为非负整数，$r = \operatorname{rank}A$.

结论 2.2.2 设 $A \in \mathbb{R}^{n\times n}$，则 $AA^T = A^TA \Leftrightarrow \exists$ 正交阵 $Q \in \mathbb{R}^{n\times n}$使得 $A = QRQ^t$，其中 $R = \begin{pmatrix} R_{11} & & \\ & \ddots & \\ & & R_{mm} \end{pmatrix}$，且 $R_{ii}(i = 1,2,\cdots,m)$或是 1 阶或是形如$\begin{pmatrix} \lambda & \mu \\ -\mu & \lambda \end{pmatrix}$的 2 阶实矩阵.

复数域上给定一个 Hermite 阵 A，通常人们称结论 1 中 k 为 A 的**正惯性指标**，$r-k$ 为 A 的**负惯性指标**，称$\{k, r-k\}$为 A 的**惯性指标**.关于 Hermite 阵的惯性指标，有著名的 Sylvester惯性定理：

定理 2.2.3 合同的 Hermite 阵，其惯性指标是相同的，从而惯性指标是 Hermite 阵的合同不变量.

证明：(1)利用二次型的基本性质来证明此结论的过程，可参见北大《高等代数》.

(2)下面利用 Schur 分解来证明之.分别让

$$B = VAV^*$$

$$A = P\begin{pmatrix} I_k & & \\ & -I_{r-k} & \\ & & 0 \end{pmatrix}P^*$$

$$B = Q\begin{pmatrix} I_{k'} & & \\ & -I_{r-k'} & \\ & & 0 \end{pmatrix}Q^*$$

其中 V, P, Q 均为可逆阵，于是

$$\begin{pmatrix} I_{k'} & & \\ & -I_{r-k'} & \\ & & 0 \end{pmatrix} = Q^{-1}VAV^*(Q^*)^{-1} = Q^{-1}VP\begin{pmatrix} I_k & & \\ & -I_{r-k} & \\ & & 0 \end{pmatrix}P^*V^*(Q^{-1})^*$$

令 $R = Q^{-1}VP$，则有

$$\begin{pmatrix} I_{k'} & & \\ & -I_{r-k'} & \\ & & 0 \end{pmatrix} = R\begin{pmatrix} I_k & & \\ & -I_{r-k} & \\ & & 0 \end{pmatrix}R^*$$

将 R^* 分块如下：$R^* = \begin{pmatrix} (R_1)_{k\times k'} & R_2 \\ R_3 & R_4 \end{pmatrix}$.若 $k' > k$，则有 R_1 的列线性相关，从而$\exists 0 \neq x_1 \in \mathbb{F}^{k'}$使得 $R_1x_1 = 0$.令 $x = \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \in \mathbb{F}^n$，又令 $R^*x = \begin{pmatrix} 0 \\ y \end{pmatrix}$，其中 $y = R_3x_1 \in \mathbb{F}^{n-k}$，于是

$$x^*\begin{pmatrix} I_{k'} & & \\ & -I_{r-k'} & \\ & & 0 \end{pmatrix}x = x^* R\begin{pmatrix} I_k & & \\ & -I_{r-k} & \\ & & 0 \end{pmatrix}R^* x$$

得 $0 < \| x_1 \|^2 = x_1^* x_1 = y^* \begin{pmatrix} -I_{r-k} & \\ & 0 \end{pmatrix} y \leqslant 0$.矛盾.类似地可证不可能有 $k > k'$,故 $k' = k$.

□

利用 Hermite 阵的惯性指标可以方便地引出下列几类特殊的 Hermite 阵:正定阵,半正定阵,负定阵,半负定阵.显然对一个 Hermite 阵 A 而言,有下列几条等价:

(1)A 是正定的(半正定的);

(2)A 的每一个特征值 $>0(\geqslant 0)$;

(3)A 合同于单位阵 $I_n\left(\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}, r = \operatorname{rank}A\right)$;

(4)A 的各阶顺序主子式 >0(所有主子式 $\geqslant 0$).

命题 2.2.1　设 A 为 n 阶 Hermite 半正定阵,则 $\forall k \in N$,存在唯一的半正定 Hermite 阵 B,使得 $A = B^k$.

证明: 只证唯一性(存在性显然).

设 $A = C^k = B^k$,B,C 为 Hermite 半正定阵.让 $C = VDV^*$,其中 V 为酉阵,D 为对角阵 $\operatorname{diag}(d_1,\cdots,d_n)$,$d_1 \geqslant \cdots \geqslant d_n \geqslant 0$,于是 $A = VD^kV^*$,$\sigma(A) = \{d_1^k,\cdots,d_n^k\}$,记 $V = (v_1,\cdots,v_n)$,于是 $\{v_1,\cdots,v_n\}$ 为对应于 $\{d_1^k,\cdots,d_n^k\}$ 的特征向量,结果

$$N(d_i^kI_n - A) = N(d_iI_n - C)$$

同样让 $B = UGU^*$,其中 U 为酉阵,G 为对角阵 $\operatorname{diag}(g_1,\cdots,g_n)$,$g_1 \geqslant \cdots \geqslant g_n \geqslant 0$,则有:

$$N(g_i^kI_n - A) = N(g_iI_n - B)$$

由 $N(d_i^kI_n - A) = N(g_i^kI_n - A)$ 得:$N(g_iI_n - B) = N(d_iI_n - C)$.注意到 $\sigma(C) = \sigma(B)$,因此,$\forall \lambda \in \sigma(C) = \sigma(B)$,均有 $N(\lambda I_n - B) = N(\lambda I_n - C)$,此即导致 $B = C$.

□

由命题 2.2.1 可知,对一个 n 阶 Hermite 半正定阵,若记 $A^{\frac{1}{k}}$ 为 A 的 k 次方根,它表示使得 $A = X^k$ 的唯一半正定阵 X,于是有 $\sqrt[k]{A}$ 与 A 可换且 $\operatorname{rank}\sqrt[k]{A} = \operatorname{rank}A$.进一步利用此事实,我们还可证得:

命题 2.2.2　设 A,B 为两个 n 阶 Hermite 阶,A 为正定阵.则 $\exists Q$ 可逆使得:

$$Q^*AQ = I_n, Q^*BQ = D$$

其中,$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$,$\{\lambda_1,\cdots,\lambda_n\} = \sigma(A^{-1}B)$,若记 $Q = (q_1,\cdots,q_n)$,则有 $Bq_j = \lambda_jAq_j$,$j = 1,2,\cdots,n$.

证明: 由 A 正定知,$(A^{\frac{1}{2}})^{-1}B(A^{\frac{1}{2}})^{-1}$ 仍为 Hermite 阵,从而存在酉阵 U 使得:

$$U^*(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})U = D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

D 为实对角阵.令 $Q = A^{-\frac{1}{2}}U$,于是 Q 可逆,它满足

$$Q^*BQ = D$$

$$Q^*AQ = U^*A^{-\frac{1}{2}}AA^{-\frac{1}{2}}U = I_n$$

再看 $\lambda_1,\cdots,\lambda_n$.由于

$$D = U^*(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}U) = U^*A^{\frac{1}{2}}(A^{-1}B)A^{-\frac{1}{2}}U = U^*A^{\frac{1}{2}}(A^{-1}B)(U^*A^{\frac{1}{2}})^{-1}$$

于是

$$\sigma(A^{-1}B) = \{\lambda_1,\cdots,\lambda_n\} = \sigma(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})$$

最后由 $AQD = BQ$ 立即知,$Bq_j = \lambda_j Aq_j$,$(j = 1,2,\cdots,n)$.

□

结论 2.2.3 设 $A\in\mathbb{C}^{n\times n}$,则 A 为正规阵$\Leftrightarrow$存在酉阵 U 使得 $A = UDU^*$,其中 D 为 $\mathbb{C}^{n\times n}$中的对角阵.

注意到酉阵为正规阵,因此由结论 2.2.3 又得:

结论 2.2.4 A 为酉阵$\Leftrightarrow A$ 为特征值均坐落在单位圆周上的正规阵.

现在由结论 2.2.3,总可以设一个正规阵 $A = U\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}U^*$,让 $U = (u_1,\cdots,u_n)$,于是 A 又可表示为 $A = \lambda_1 u_1 u_1^* + \cdots + \lambda_n u_n u_n^*$.记 $A_i = u_i u_i^*$,则$\{A_1,\cdots,A_n\}$满足

$$\begin{cases} A_i^* = A_i \neq 0 \ (i = 1,2,\cdots,n) \\ A_iA_j = 0 \ (i \neq j) \\ \sum\limits_{i=1}^{n} A_i = I \end{cases} \tag{2.2.1}$$

一般地,若$\{A_1,\cdots,A_n\}$满足(2.2.1)式,而 $A = \sum\limits_{i=1}^{n}\lambda_i A_i$,则称 $A = \sum\limits_{i=1}^{n}\lambda_i A_i$ 为正规阵 A 的**谱分解**.容易知道,给定 $A\in\mathbb{C}^{n\times n}$,则有以下两条等价:

(1)A 为正规阵;

(2)A 存在谱分解 $A = \sum\limits_{i=1}^{n}\lambda_i A_i$,其中$\{A_1,\cdots,A_n\}$满足(2.2.1)式.

因此,关于矩阵的谱分解实际上只对正规矩阵有定义.换句话说,不可酉对角化的矩阵没有谱分解.

定理 2.2.4 设 $A,B\in\mathbb{C}^{n\times n}$,$A$ 与 B 都是正规阵,则有以下两条等价.

(1)$AB = BA$;

(2)存在酉阵 U 使得 UAU^* 与 UBU^* 均为对角阵.

证明:(2)$\Rightarrow$(1) 显然.

(1)⇒(2)　A 与 B 都是正规阵，于是可设 A 的谱分解形式 $A = \sum \lambda_i E_i$，其中 λ_i 各不相同；又设 B 的谱分解形式为 $B = \sum \mu_j F_j$，其中 μ_j 各不相同，则由 $AB = BA$ 导致任何 E_i 与任何 F_j 皆可换(关于此论断的证明作为练习).结果：

$$A = \sum \lambda_i E_i \sum F_j = \sum \lambda_i E_i F_j$$

$$B = \sum \mu_j E_i F_j$$

故 A 与 B 可以同时酉相似对角化.

□

结论 2.2.5　设 $A \in \mathbb{C}^{n\times n}$，$\mathrm{spec}A = \{\lambda_1, \cdots, \lambda_n\}$，则有 Schur 不等式成立：

$$\sum_{i=1}^{n} |\lambda_i|^2 \leqslant \sum_{i=1}^{n} |a_{ij}|^2$$

进一步，该不等式等号成立当且仅当 A 为正规矩阵.

证明： 取 A 的 Schur 分解：$UAU^* = \begin{pmatrix} \lambda_1 & * & * \\ & \ddots & \vdots \\ & & \lambda_n \end{pmatrix}$.于是 $UA^*U^* = \begin{pmatrix} \bar{\lambda}_1 & & \\ \vdots & \ddots & \\ * & \cdots & \bar{\lambda}_n \end{pmatrix}$.

结果，$UAA^*U^* = \begin{pmatrix} c_{11} & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & c_{nn} \end{pmatrix}$，两边取迹得 $\sum_{i,j} |a_{ij}|^2 \geqslant \sum_{i=1}^{n} |\lambda_i|^2$，并且等号成立当且仅当 A 为正规矩阵.

□

Schur 分解还可以用来对矩阵的秩进行估计.任取 $M \in \mathbb{C}^{n\times n}$，$1 \leqslant k \leqslant n-1$，将 M 分块为 $M = \begin{pmatrix} \overset{k}{\frown} & \overset{n-k}{\frown} \\ A & B \\ C & D \end{pmatrix}$，记 $l_k = \mathrm{tr}(AA^*) + \mathrm{tr}(DD^*) + 2\sqrt{\mathrm{tr}(BB^*)\mathrm{tr}(CC^*)}$，规定 $l_n = \mathrm{tr}(MM^*)$，$l = \min\limits_{1\leqslant k\leqslant n} l_k$，则有如下结论：

定理 2.2.5　M, l 如上所设，则有 $l\mathrm{rank}(M) \geqslant |\mathrm{tr}(M)|^2$.

证明： 可设 $M \neq 0$，取 M 的 Schur 分解：$UMU^* = \begin{pmatrix} \lambda_1 & \cdots & * \\ & \ddots & \vdots \\ & & \lambda_n \end{pmatrix}$.设 M 共有 s 个非零特征值，则 $s \leqslant \mathrm{rank}(M)$.不失一般性，令 $\lambda_1, \cdots, \lambda_s$ 为非零特征值，于是

$$|\mathrm{tr}(M)|^2 = \left|\sum_{i=1}^{s} \lambda_i\right|^2 \leqslant s\sum_{i=1}^{s} |\lambda_i|^2 \leqslant \mathrm{rank}(M)\sum_{i=1}^{s} |\lambda_i|^2$$

不妨设 $B \neq 0, C \neq 0$.取 $\zeta_1^2 = \mathrm{tr}(BB^*) > 0$，$\zeta_2^2 = \mathrm{tr}(CC^*) > 0$，构造方阵 $K = \begin{pmatrix} A & \sqrt{\dfrac{\zeta_2}{\zeta_1}}B \\ \sqrt{\dfrac{\zeta_1}{\zeta_2}}C & D \end{pmatrix}$，则有 $KK^* = \begin{pmatrix} AA^* + \dfrac{\zeta_2}{\zeta_1}BB^* & * \\ * & \dfrac{\zeta_1}{\zeta_2}CC^* + DD^* \end{pmatrix}$.结果：

$$\mathrm{tr}(KK^*)=\mathrm{tr}(AA^*)+\mathrm{tr}(DD^*)+2\zeta_1\zeta_2=l_k$$

又因为有等式

$$K=\begin{pmatrix}\sqrt{\frac{\zeta_2}{\zeta_1}}I_k & 0\\ 0 & I_{n-k}\end{pmatrix}\begin{pmatrix}A & B\\ C & D\end{pmatrix}\begin{pmatrix}\sqrt{\frac{\zeta_1}{\zeta_2}}I_k & 0\\ 0 & I_{n-k}\end{pmatrix}$$

因此 $K\sim M,\lambda_1,\cdots,\lambda_n$ 也为 K 的特征值，从而据 Schur 不等式：

$$\sum_{i=1}^{s}|\lambda_i|^2\leqslant\mathrm{tr}(KK^*)$$

得

$$|\mathrm{tr}(M)|^2\leqslant\mathrm{rank}(M)\mathrm{tr}(KK^*)=l_k\cdot\mathrm{rank}(M)$$

容易知道，此不等式对 $B=0$ 或 $C=0$ 时仍然成立（此时令 $K=M$）.所以

$$|\mathrm{tr}(M)|^2\leqslant l\cdot\mathrm{rank}(M)$$

□

由定理 2.2.4 易知以下推论成立：

推论 2.2.1 M,l 与定理 2.2.5 相同，如果 $|\mathrm{tr}(M)|^2>(n-1)l$，则 M 可逆.

证明： $\mathrm{rank}(M)\geqslant\frac{1}{l}|\mathrm{tr}(M)|^2>\frac{1}{l}(n-1)l=n-1$，故 M 可逆.

□

推论 2.2.2 任取 $0\neq A\in\mathbb{C}^{m\times n}$，则有 $\mathrm{rank}(A)\geqslant\frac{[\mathrm{tr}(AA^*)]^2}{\mathrm{tr}(AA^*)^2}$.

证明： 设 M 为非零方阵，易知有下列两式成立：

$$|\mathrm{tr}(M)|^2\leqslant\mathrm{rank}(M)\sum_{i=1}^{n}|\lambda_i|^2,\ \sum_{i=1}^{n}|\lambda_i|^2\leqslant\mathrm{tr}(MM^*)$$

从而 $\mathrm{rank}(M)\geqslant\frac{|\mathrm{tr}(M)|^2}{\mathrm{tr}(MM^*)}$.

取 $M=AA^*$，则得 $\mathrm{rank}(A)=\mathrm{rank}(M)\geqslant[\mathrm{tr}(AA^*)]^2/\mathrm{tr}((AA^*)^2)$.

□

推论 2.2.3 设 $A\in\mathbb{C}^{m\times n}$，若 $[\mathrm{tr}(A^*A)]^2>(n-1)\mathrm{tr}((A^*A)^2)$，则 A 为列满秩阵.

证明： 显然.

□

2.3 矩阵的奇异值分解

我们知道，复数域 $\mathbb{F}$ 上两个 $m\times n$ 矩阵 A 和 B 称为**相抵**（或**等价**）是指，存在可逆 $P\in\mathbb{F}^{m\times m}$ 及可逆 $Q\in\mathbb{F}^{n\times n}$，使得 $A=PBQ$，常记作 $A\approx B$. 相抵关系是等价关系. $\forall A\in\mathbb{F}^{m\times n}$，记 $r=\mathrm{rank}A$，则必有 $A\approx\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}$，它为相抵标准型. 下面介绍一种很特殊的相抵

——酉相抵.

定义 2.3.1 $\mathbb{F}$ 上 $m\times n$ 矩阵A 和B 称为**酉相抵**,如果有 m 阶和n 阶酉方阵U 和V,使 $UAV=B$,常记作 $A \overset{酉}{\approx} B$.

显然,酉相抵关系也是等价关系.那么能否像相抵等价类的标准型一样,找出酉相抵等价类的标准型呢?结论是肯定的.

定理 2.3.1(酉相抵标准型定理) 设 $A\in\mathbb{F}^{m\times n}$,$\mathrm{rank}A=r$,记 $\mathrm{spec}A^*A=\{\sigma_1^2,\cdots,\sigma_r^2,0,\cdots,0\}$,$\sigma_1,\cdots,\sigma_r>0$,则 A 酉相抵于矩阵$\begin{pmatrix} D_1 & 0\\ 0 & 0\end{pmatrix}_{m\times n}$,其中 $D_1=\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}_{r\times r}$.

证明:注意到 A^*A 为半正定矩阵,记其特征值为 $\sigma_1^2,\cdots,\sigma_r^2,0,\cdots,0$.

其对应的标准正交特征向量是 $x_1,\cdots,x_r,x_{r+1},\cdots,x_n$.由内积定义知:

$$\langle Ax_i,Ax_j\rangle=\langle x_i,A^*Ax_j\rangle=\sigma_j^2\langle x_i,x_j\rangle=\begin{cases}\sigma_j^2, & i=j\text{ 时}\\ 0, & i\neq j\text{ 时}\end{cases}$$

因此 $Ax_1,\cdots,Ax_r$ 仍是彼此正交的非零向量.记

$$V=(x_1,\cdots,x_n)$$

$$U=\left(\frac{Ax_1}{\sigma_1},\cdots,\frac{Ax_r}{\sigma_r},y_{r+1},\cdots,y_m\right)$$

其中 $y_{r+1},\cdots,y_m$ 通过将$\frac{Ax_1}{\sigma_1},\cdots,\frac{Ax_r}{\sigma_r}$扩充为$\mathbb{F}^m$ 的标正基而得到.于是

$$\begin{aligned}U^*AV&=U^*(Ax_1,\cdots,Ax_r,0,\cdots,0)\\ &=\begin{pmatrix}\left(\frac{Ax_1}{\sigma_1}\right)^*\\ \vdots\\ \left(\frac{Ax_r}{\sigma_r}\right)^*\\ \vdots\end{pmatrix}(Ax_1,\cdots,Ax_r,0,\cdots,0)=\begin{pmatrix}\sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0\end{pmatrix}_{m\times n}\end{aligned}$$

定理获证.

□

定义 2.3.2 设 $A\in\mathbb{F}^{m\times n}$,记 $\mathrm{spec}A^*A=\{\sigma_1^2,\cdots,\sigma_n^2\}$,$\sigma_1\geqslant\sigma_2\cdots\geqslant\sigma_n\geqslant 0$,则称 $\sigma_1,\cdots,\sigma_n$ 中的非零元$\sigma_1,\cdots,\sigma_r$ 为A 的**非零奇异值**.

注意到 A^*A 与 AA^* 的非零特征值总是一致的,因此若记 $\mathrm{spec}AA^*=\{s_1^2,\cdots,s_n^2\}$,其中 $s_1\geqslant s_2\cdots\geqslant s_r>s_{r+1}=\cdots=s_m=0$,则有 $s_1,s_2,\cdots,s_r$ 即为A 的非零奇异值.换句话说,A 的非零奇异值的定义也可通过 $\mathrm{spec}AA^*$ 来定义.一般地,通过 A^*A,人们定义 $\sigma_1,\cdots,\sigma_n$ 为 A 的奇异值,而通过 AA^*,人们又可定义 $s_1,s_2,\cdots,s_m$ 为A 的奇异值,但这两种定义却未必一致,因为未必有 $m=n$.

由定理 2.3.1 知,$\forall A\in\mathbb{F}^{m\times n}$,总可取 A 的如下分解:

$$A=U\Sigma_AV^*$$

其中 U 为 m 阶酉阵，V 为 n 阶酉阵，$\Sigma_A=\begin{pmatrix}\sigma_1&&&\\&\ddots&&\\&&\sigma_r&\\&&&0\end{pmatrix}_{m\times n}$，$r=\mathrm{rank}A$，$\sigma_1\geqslant\sigma_2\cdots\geqslant\sigma_r>0$ 为 A 的非零奇异值.通常称该分解为 A 的**奇异值分解**.此时 $A\overset{\text{酉}}{\approx}\Sigma_A$.进一步，通过下列命题2.3.1，证明 A 的上述奇异值分解中的 Σ_A 还可以充当 A 所在的酉相抵等价类的标准型.

命题 2.3.1 设 $A,B\in\mathbb{F}^{m\times n}$，则 $A\overset{\text{酉}}{\approx}B\Leftrightarrow \mathrm{spec}AA^*=\mathrm{spec}BB^*\Leftrightarrow A$ 与 B 的非零奇异值相同.

证明： $A\overset{\text{酉}}{\approx}B$ 时，显然有 AA^* 相似于 BB^*，因此 $\mathrm{spec}AA^*=\mathrm{spec}BB^*$；反之，若 $\mathrm{spec}AA^*=\mathrm{spec}BB^*$，则有 $A\overset{\text{酉}}{\approx}\Sigma_A=\Sigma_B\overset{\text{酉}}{\approx}B$.

□

将 A 的奇异值分解 $A=U\Sigma_AV^*$ 进行移项得：

$$AV=U\begin{pmatrix}\sigma_1&&&\\&\ddots&&\\&&\sigma_r&\\&&&0\end{pmatrix},U^*A=\begin{pmatrix}\sigma_1&&&\\&\ddots&&\\&&\sigma_r&\\&&&0\end{pmatrix}V^*$$

若记 $U=(u_1,\cdots,u_m)$，$V=(v_1,\cdots,v_n)$，则有

$$\begin{cases}Av_1=u_1\sigma_1\\ \quad\vdots\\ Av_r=u_r\sigma_r\end{cases}\qquad\begin{cases}u_1^*A=\sigma_1v_1^*\\ \quad\vdots\\ u_r^*A=\sigma_rv_r^*\end{cases}$$

一般地，称 $v_1,\cdots,v_r$ 为 A 关于奇异值 $\sigma_1,\cdots,\sigma_r$ 对应的**右奇异向量**，而 $u_1^*,\cdots,u_r^*$ 称为 A 关于奇异值 $\sigma_1,\cdots,\sigma_r$ 对应的**左奇异向量**.

矩阵的奇异值分解有着广泛的应用，下面的几个结论可看成其众多应用中的一小部分.

定理 2.3.2 设 $A\in\mathbb{F}^{m\times n}(m\leqslant n)$，则矩阵 A 可以分解为 $A=HU$，其中 $H_{m\times m}$ 为半正定阵且与 A 有相同的非零奇异值，而 $U\in\mathbb{F}^{m\times n}$ 满足 $UU^*=I_m$.进一步，H 是唯一的，它是 $\sqrt{AA^*}$.

证明： 取 A 的奇异值分解 $A=U_1\begin{pmatrix}D&0\\0&0\end{pmatrix}V^*$，其中 $D=\begin{pmatrix}\sigma_1&&\\&\ddots&\\&&\sigma_r\end{pmatrix}$ 为对角阵，$\sigma_1\geqslant\sigma_2\cdots\geqslant\sigma_r>0$.将 V 分块为 $V=(V_1,V_2)$，其中 $V_1\in\mathbb{F}^{n\times m}$，$V_2\in\mathbb{F}^{n\times(n-m)}$，将 $\begin{pmatrix}D&0\\0&0\end{pmatrix}$ 分块为 $(\overset{m}{\Sigma}\quad\overset{n-m}{0})$，于是

$$A=U_1(\Sigma,0)\begin{pmatrix}V_1^*\\V_2^*\end{pmatrix}=(U_1\Sigma,0)\begin{pmatrix}V_1^*\\V_2^*\end{pmatrix}$$

$$= U_1 \Sigma V_1^* = U_1 \Sigma U_1^* \cdot U_1 V_1^*$$

令 $H = U_1 \Sigma U_1^*$，为 m 阶半正定阵，$U = U_1 V_1^*$ 满足 $UU^* = I_m$. 于是定理的存在性获证.

进一步，若 A 还有分解 $A = H_2 U_2$，其中 H_2 为 m 阶半正定且与 A 有相同的非零奇异值，而 $U_2 \in \mathbb{F}^{m\times n}$ 是满足 $U_2 U_2^* = I_m$ 的部分酉阵，则有：

$$AA^* = H_2 U_2 U_2^* H_2^* = H_2^2,\ H_2 = \sqrt{AA^*}，唯一.$$

□

类似地，还可得：$\forall A \in \mathbb{F}^{m\times n}(m \geqslant n)$，$A$ 必有分解 $A = UH$，其中 H 为 n 阶半正定阵且与 A 有相同的非零奇异值，而 $U \in \mathbb{F}^{m\times n}$ 且满足 $U^* U = I_n$，并且 H 是唯一的，它为 $\sqrt{A^* A}$.

通常，人们把定理 2.3.2 中的 A 分解叫做 A 的**极分解**. 这是因为，若将 $\mathbb{F}^{n\times n}$ 与 $\mathbb{C}$ 作类比，半正定 Hermite 阵恰可看成矩阵中的"非负数"，而酉阵则可看成矩阵中的"模为 1 的复数".

定理 2.3.3（$C-S$ 分解） 设酉阵 $W \in \mathbb{C}^{n\times n}$ 分块为

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix} \begin{matrix} l \\ n-l \end{matrix}$$
$$\begin{matrix} l & n-l \end{matrix}$$

其中 $2l \leqslant n$，则存在酉阵 $U = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix}$ 及酉阵 $V = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$，其中 U_1, V_1 为 l 阶方阵，U_2, V_2 为 $n-l$ 阶方阵，使得

$$U^* WV = \begin{pmatrix} C & -S & 0 \\ S & C & 0 \\ 0 & 0 & I \end{pmatrix} \begin{matrix} l \\ l \\ n-2l \end{matrix}$$

其中

$$C = \begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_l \end{pmatrix}, S = \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_l \end{pmatrix}, c_i = \cos\theta_i, s_i = \sin\theta_i, 0 \leqslant \theta_l \leqslant \theta_{l-1} \leqslant \cdots \leqslant \theta_1 \leqslant \frac{\pi}{2}.$$

证明： 由 W 为酉阵知：

$$\begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix}^* \begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} = I_l$$

因此有

$$W_{11}^* W_{11} = I_l - W_{21}^* W_{21}$$

从而 $W_{11}^* W_{11}$ 的特征值均为不大于 1 的非负数. 若 W_{11} 的奇异值分解为

$$U_1^* W_{11} V_1 = C$$

其中 U_1, V_1 皆为 l 阶酉阵，$C = \begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_l \end{pmatrix}$ 且 $0 \leqslant c_1 \leqslant \cdots \leqslant c_k < c_{k+1} = \cdots = c_l = 1$. 注意

到

$$\begin{aligned} I_l &= V_1^*(W_{11}^*, W_{21}^*)\begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} V_1 \\ &= (V_1^* W_{11}^*, V_1^* W_{21}^*)\begin{pmatrix} W_{11} V_1 \\ W_{21} V_1 \end{pmatrix} \\ &= V_1^* W_{11}^* W_{11} V_1 + V_1^* W_{21}^* W_{21} V_1 \\ &= C^2 + (W_{21} V_1)^* W_{21} V_1 \end{aligned}$$

故有

$$(W_{21} V_1)^* W_{21} V_1 = I_l - C^2 = \begin{pmatrix} 1-c_1^2 & & & & & \\ & \ddots & & & & \\ & & 1-c_k^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

因此知：$W_{21} V_1$ 的前 k 列相互正交，后 $l-k$ 列均为 0，现将 $W_{21} V_1$ 的前 k 列单位化后扩充为列向量空间$\mathbb{C}^{(n-l)}$的标准正交基，从而获得 $n-l$ 阶酉阵 $\widetilde{U}_2$，此时有

$$\widetilde{U}_2^* W_{21} V_1 = \begin{pmatrix} S \\ 0 \end{pmatrix}$$

其中 $S = \begin{pmatrix} \sqrt{1-c_1^2} & & & & & \\ & \ddots & & & & \\ & & \sqrt{1-c_k^2} & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}_{l \times l}$. 自然 $S^2 = I_l - C^2$.

结合 $U_1^* W_{11} V_1 = C$ 又知

$$\begin{pmatrix} U_1^* & 0 \\ 0 & \widetilde{U}_2^* \end{pmatrix}\begin{pmatrix} W_{11} \\ W_{21} \end{pmatrix} V_1 = \begin{pmatrix} C \\ S \\ 0 \end{pmatrix} \begin{matrix} l \\ l \\ n-2l \end{matrix}$$

现考虑 W 的行(W_{11}, W_{12})，由

$$\begin{aligned} I_l &= U_1^*(W_{11}, W_{12})\begin{pmatrix} W_{11}^* \\ W_{12}^* \end{pmatrix} U_1 \\ &= (U_1^* W_{11}, U_1^* W_{12})\begin{pmatrix} W_{11}^* U_1 \\ W_{12}^* U_1 \end{pmatrix} \\ &= C^2 + (U_1^* W_{12}) W_{12}^* U_1 \end{aligned}$$

得

$$(U_1^* W_{12}) W_{12}^* U_1 = I_l - C^2 = \begin{pmatrix} 1-c_1^2 & & & & & \\ & \ddots & & & & \\ & & 1-c_k^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

因此 $U_1^* W_{12}$ 的前 k 行相互正交，后 $l-k$ 行全为 0. 将 $U_1^* W_{12}$ 的前 k 行分别乘以 $-\frac{1}{\sqrt{1-c_i^2}}(i=1,2,\cdots,k)$，再将之扩充为行向量空间 $\mathbb{C}^{n-l}$ 的标正基 $\{\beta_1,\cdots,\beta_{n-l}\}$，将酉阵 $\begin{pmatrix}\beta_1\\ \vdots \\ \beta_{n-l}\end{pmatrix}$ 通过转置共轭得酉阵 V_2，于是 V_2 满足 $U_1^* W_{12} V_2 = (-S,0)$. 结果 $\begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}$ 仍为 n 阶酉阵且

$$\begin{matrix} l \\ n-l \end{matrix}\begin{pmatrix} U_1^* & 0 \\ 0 & \tilde{U}_2^* \end{pmatrix}\begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}\begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}\begin{matrix} l \\ n-l \end{matrix}$$

$$\begin{pmatrix} U_1^* W_{11} V_1 & U_1^* W_{12} V_2 \\ \tilde{U}_2^* W_{21} V_1 & \tilde{U}_2^* W_{22} V_2 \end{pmatrix} = \begin{pmatrix} C & -S & 0 \\ S & \square & \square \\ 0 & \square & \square \end{pmatrix}$$

其中，$\tilde{U}_2^* W_{22} V_2 = \begin{pmatrix} \square & \square \\ \square & \square \end{pmatrix}$. 令 $C_1 = \begin{pmatrix} c_1 & & \\ & \ddots & \\ & & c_k \end{pmatrix}$，又有

$$\begin{pmatrix} C & -S & 0 \\ S & \square & \square \\ 0 & \square & \square \end{pmatrix} = \begin{pmatrix} C_1 & 0 & -S_1 & 0 & 0 \\ 0 & I_{l-k} & 0 & 0 & 0 \\ S_1 & 0 & X_{33} & X_{34} & X_{35} \\ 0 & 0 & X_{43} & X_{44} & X_{45} \\ 0 & 0 & X_{53} & X_{54} & X_{55} \end{pmatrix}$$

其中 $S_1 = \begin{pmatrix} \sqrt{1-c_1^2} & & \\ & \ddots & \\ & & \sqrt{1-c_k^2} \end{pmatrix}_{k\times k}$，$S_1$ 可逆．由酉阵的定义立即知：

$$X_{33} = C_1, X_{34} = 0 = X_{35}, X_{43} = 0 = X_{53}$$

因此

$$\begin{pmatrix} U_1^* & 0 \\ 0 & \tilde{U}_2^* \end{pmatrix} W \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} = \begin{pmatrix} C_1 & 0 & -S_1 & 0 & 0 \\ 0 & I_{l-k} & 0 & 0 & 0 \\ S_1 & 0 & C_1 & 0 & 0 \\ 0 & 0 & 0 & X_{44} & X_{45} \\ 0 & 0 & 0 & X_{54} & X_{55} \end{pmatrix}$$

其中$\begin{pmatrix} X_{44} & X_{45} \\ X_{54} & X_{55} \end{pmatrix}$仍为酉阵.再令 $U_3 = \begin{pmatrix} X_{44} & X_{45} \\ X_{54} & X_{55} \end{pmatrix}$,于是$\begin{pmatrix} I_{l+k} & 0 \\ 0 & U_3 \end{pmatrix}$仍为酉阵,且

$$\begin{pmatrix} I_{l+k} & 0 \\ 0 & U_3^* \end{pmatrix}\begin{pmatrix} U_1^* & 0 \\ 0 & \widetilde{U}_2^* \end{pmatrix} W \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} = \begin{pmatrix} C_1 & 0 & -S_1 & 0 & 0 \\ 0 & I_{l-k} & 0 & 0 & 0 \\ S_1 & 0 & C_1 & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix}$$

注意到

$$\begin{pmatrix} I_{l+k} & 0 \\ 0 & U_3^* \end{pmatrix}\begin{pmatrix} U_1^* & 0 \\ 0 & \widetilde{U}_2^* \end{pmatrix} = \begin{pmatrix} I_l & & \\ & I_k & \\ & & U_3^* \end{pmatrix}\begin{pmatrix} U_1^* & 0 \\ 0 & \widetilde{U}_2^* \end{pmatrix} = \begin{pmatrix} U_1^* & 0 \\ 0 & U_2^* \end{pmatrix}$$

其中 $U_2^* = \begin{pmatrix} I_k & 0 \\ 0 & U_3^* \end{pmatrix}\widetilde{U}_2^*$ 为酉阵.结果

$$\begin{pmatrix} U_1^* & 0 \\ 0 & U_2^* \end{pmatrix} W \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} = \begin{pmatrix} C_1 & 0 & -S_1 & 0 & 0 \\ 0 & I_{l-k} & 0 & 0 & 0 \\ S_1 & 0 & C_1 & 0 & 0 \\ 0 & 0 & 0 & I_{l-k} & 0 \\ 0 & 0 & 0 & 0 & I_{n-2l} \end{pmatrix} = \begin{pmatrix} C & -S & 0 \\ S & C & 0 \\ 0 & 0 & I \end{pmatrix}$$

定理获证.

□

作为奇异值分解的另外一个应用,本节最后给出两个同型矩阵能同时酉相抵于对角形的充要条件.

定理 2.3.4 设 $A,B\in \mathbb{F}^{m\times n}$,则 A,B 可同时酉相抵于对角阵的充要条件是 AB^* 与 B^*A 均为正规阵.

为证此定理,首先证明如下引理:

引理 2.3.1 设矩阵 $A = \begin{pmatrix} \mu_1 I_{n_1} & & & \\ & \ddots & & \\ & & \mu_t I_{n_t} & \\ & & & 0 \end{pmatrix} \in \mathbb{F}^{m\times n}$, $n_1 + \cdots + n_t = r$, $|\mu_1|,\cdots,|\mu_t|$为各不相同的正数.若 $B\in \mathbb{F}^{m\times n}$,且 AB^* 和 B^*A 均为正规阵,则 B 必形如 $\begin{pmatrix} B_1 & & & \\ & \ddots & & \\ & & B_t & \\ & & & B_{t+1} \end{pmatrix}$,其中 $B_i\in \mathbb{F}^{n_i\times n_i}$,$(i=1,2,\cdots,t)$均为正规阵.

证明: 将 B 作形同A 的分块,其(i,j)位置上的块元为 B_{ij},于是

$$A^* BB^* A = B^* AA^* B$$
$$AB^* BA^* = BA^* AB^*$$

计算上面两式中的第 i 对角块得($i=1,2,\cdots,t$):

$$|\mu_i|^2 \sum_k B_{ik}B_{ik}^* = \sum_k |\mu_k|^2 B_{ki}^* B_{ki} \tag{2.3.1}$$

$$|\mu_i|^2 \sum_k B_{ki}^* B_{ki} = \sum_k |\mu_k|^2 B_{ik}B_{ik}^* \tag{2.3.2}$$

由上两式相减得 $\sum_k (|\mu_i|^2 - |\mu_k|^2) B_{ik}B_{ik}^* = \sum_k (|\mu_k|^2 - |\mu_i|^2) B_{ki}^* B_{ki}$,从而

$$\sum_k (|\mu_k|^2 - |\mu_i|^2)(B_{ik}B_{ik}^* + B_{ki}^* B_{ki}) = 0 \tag{2.3.3}$$

由于 $B_{ik}B_{ik}^*$ 与 $B_{ki}^*B_{ki}$ 均为半正定阵,且不妨设 $|\mu_1| = \max\{|\mu_1|,\cdots,|\mu_t|\}$,则有

$$B_{1j} = 0\ (j=2,3,\cdots)$$
$$B_{l1} = 0\ (l=2,3,\cdots)$$

类似地,从(2.3.3)式进一步可得 $B_{ij}=0, i\neq j, i=1,2,\cdots,t,t+1, j=1,2,\cdots,t,t+1$.
因此有

$$B = \begin{pmatrix} B_{11} & & & \\ & \ddots & & \\ & & B_{tt} & \\ & & & B_{t+1} \end{pmatrix}$$

最后 AB^* 与 B^*A 均为正规阵蕴含 $B_{11},\cdots,B_{tt}$ 均为正规阵,引理获证.

□

定理 2.3.4 的证明　⇒　由已知∃ U,V 皆为酉阵使得 U^*AV 与 U^*BV 皆为对角阵,因此,$U^*AB^*U = U^*AV(U^*BV)^*$ 为对角阵,故 AB^* 为正规阵,同理知,BA^* 为正规阵.

⇐　取 A 的奇异值分解 $\Sigma_A = U_1^* AV_1$, $\Sigma_A = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}$, $\sigma_1 \geqslant \sigma_2 \cdots \geqslant \sigma_r > 0$.考察 Σ_A 与 $U_1^* BV_1$,注意到:

$$AB^* = U_1 \Sigma_A V_1^* B^* = U_1(\Sigma_A (U_1^* BV_1)^*) U_1^*$$
$$BA^* = V_1((U_1^* BV_1)^* \Sigma_A) V_1^*$$

因此 $\Sigma_A (U_1^* BV_1)^*$ 和 $(U_1^* BV_1)^* \Sigma_A$ 皆为正规阵.将 Σ_A 写成

$$\Sigma_A = \begin{pmatrix} \mu_1 I_{n_1} & & & \\ & \ddots & & \\ & & \mu_t I_{n_t} & \\ & & & 0 \end{pmatrix}$$

于是利用引理 2.3.1 得

$$U_1^* BV_1 = \begin{pmatrix} B_1 & & & \\ & \ddots & & \\ & & B_t & \\ & & & B_{t+1} \end{pmatrix}$$

其中 $B_1,\cdots,B_t$ 皆为正规阵，它们均酉相似于对角阵，即分别存在酉阵 $Q_1,\cdots,Q_t$ 使得

$$Q_1^* B_1 Q_1 = \Lambda_1$$
$$\vdots$$
$$Q_t^* B_t Q_t = \Lambda_t$$

另外，取 B_{t+1} 的奇异值分解 $Q_{t+1}^* B_{t+1} V_{t+1} = \Sigma_{B_{t+1}}$，则有

$$\begin{pmatrix} Q_1^* & & & \\ & \ddots & & \\ & & Q_t^* & \\ & & & Q_{t+1}^* \end{pmatrix} U_1^* BV_1 \begin{pmatrix} Q_1 & & & \\ & \ddots & & \\ & & Q_t & \\ & & & V_{t+1} \end{pmatrix} = \begin{pmatrix} \Lambda_1 & & & \\ & \ddots & & \\ & & \Lambda_t & \\ & & & \Sigma_{B_{t+1}} \end{pmatrix}$$

$$\begin{pmatrix} Q_1^* & & & \\ & \ddots & & \\ & & Q_t^* & \\ & & & Q_{t+1}^* \end{pmatrix} U_1^* AV_1 \begin{pmatrix} Q_1 & & & \\ & \ddots & & \\ & & Q_t & \\ & & & V_{t+1} \end{pmatrix} = \begin{pmatrix} \mu_1 I_{n_1} & & & \\ & \ddots & & \\ & & \mu_t I_{n_t} & \\ & & & 0 \end{pmatrix}$$

定理获证.

□

由于 A 的奇异值的计算实质上是自伴阵 $A^* A$ 的特征值计算问题，因此它也可以有类似于特征值的变分描述.对此，将在本书最后一章通过对特征值的变分描述(Courant - Fisher 定理)来加以解决.

2.4 矩阵的 LU 分解

矩阵的 LU 分解是一类特殊的满秩分解，其定义如下：

定义 2.4.1 设 $A\in \mathbb{F}^{m\times n}$，若 A 能写成 $A=LU$，其中

L 形如 $\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ \vdots & & l_{r+1,r} \\ & & \vdots \\ l_{m1} & \cdots & l_{mr} \end{pmatrix}_{m\times r}$，$U$ 形如 $\begin{pmatrix} u_{11} & u_{12} & \cdots & & u_{1n} \\ & \ddots & & & \\ & & u_{rr} & \cdots & u_{rn} \end{pmatrix}_{r\times n}$，且 $u_{11},\cdots,u_{rr}$ 皆不为 0，则称 $A=LU$ 为 A 的 **LU 分解**.

给定矩阵 $A\in \mathbb{F}^{m\times n}$，若 $A=\tilde{L}\tilde{U}$ 为 A 的 LU 分解，将 $\tilde{U}$ 写成

$$\begin{pmatrix} u_{11} & & \\ & \ddots & \\ & & u_{rr} \end{pmatrix}\begin{pmatrix} 1 & \frac{u_{12}}{u_{11}} & \cdots & \cdots & \frac{u_{1n}}{u_{11}} \\ & \ddots & & & \vdots \\ & & 1 & \cdots & \frac{u_{rn}}{u_{rr}} \end{pmatrix}$$

记 $D=\begin{pmatrix} u_{11} & & \\ & \ddots & \\ & & u_{rr} \end{pmatrix}$, $U=\begin{pmatrix} 1 & \frac{u_{12}}{u_{11}} & \cdots & \cdots & \frac{u_{1n}}{u_{11}} \\ & u_{11} & & & u_{11} \\ & \ddots & & & \vdots \\ & & 1 & \cdots & \frac{u_{rn}}{u_{rr}} \end{pmatrix}$,则称 $A=LDU$ 为 A 的 **LDU 分解**.

下面首先解决矩阵的 LU 分解存在性问题.由 LDU 分解的定义可知 A 的 LDU 分解的存在唯一性等价于 A 的 LU 分解的存在唯一性.

定理 2.4.1　设 $A\in \mathbb{F}_r^{m\times n}$,则以下两条等价:

(1)A 有 LU 分解;

(2)A 的前 r 个顺序主子式皆不为 0.

进一步,若 $A=LU$ 及 $A=L'U'$ 皆为 A 的 LU 的分解,则必有 $L=L'$, $U=U'$.

证明: (2)⇒(1)　由 $A\begin{pmatrix}1\\1\end{pmatrix}=a_{11}\neq 0$ 知,A 通过左乘适当消元矩阵:

$$L_1=E(l_1,e_1;1)\triangleq I-l_1e_1^{\mathrm{T}}=\begin{pmatrix} 1 & & & \\ -l_{21} & 1 & & \\ \vdots & \vdots & \ddots & \\ -l_{m1} & 0 & \cdots & 1 \end{pmatrix},$$

其中 $l_1=(0,l_{21},\cdots,l_{m1})^{\mathrm{T}}$, $e_1=(1,0,\cdots,0)^{\mathrm{T}}$,使得:

$$A^{(1)}\triangleq L_1A=\begin{pmatrix} a_{11}^{(0)} & a_{12}^{(0)} & \cdots & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ \vdots & \vdots & & \vdots \\ 0 & a_{m2}^{(1)} & \cdots & a_{mn}^{(1)} \end{pmatrix}$$

这里记 $A=A^{(0)}=\begin{pmatrix} a_{11}^{(0)} & \cdots & a_{1n}^{(0)} \\ \vdots & & \vdots \\ a_{m1}^{(0)} & \cdots & a_{mn}^{(0)} \end{pmatrix}$.由于第二类行初等变换不改变行列式的值,因此 $A^{(1)}$ 的各级顺序主子式与 A 的各级顺序主子式相同.从而 $a_{22}^{(1)}\neq 0$.对 $\begin{pmatrix} a_{22}^{(1)} & \cdots & a_{2n}^{(1)} \\ \vdots & & \vdots \\ a_{m1}^{(1)} & \cdots & a_{mn}^{(1)} \end{pmatrix}$ 施行前述过程;由此重复下去得 $A=LU$,其中

$$L=\begin{pmatrix}1 & & & & & & \\ l_{21} & 1 & & & & & \\ & & \ddots & & & & \\ \vdots & & & 1 & & & \\ \vdots & & & l_{mr} & 1 & & \\ & & & \vdots & \vdots & \ddots & \\ l_{m1} & \cdots & \cdots & l_{mr} & 0 & \cdots & 1\end{pmatrix}$$

$$U=\begin{pmatrix}a_{11}^{(0)} & \cdots & \cdots & \cdots & a_{1n}^{(0)} \\ & \ddots & & & \\ & & a_{r-1,r-1}^{(r-2)} & \cdots & a_{r-1,n}^{(r-2)} \\ & & 0 & a_{rr}^{(r-1)} & a_{r,n}^{(r-1)} \\ & & 0 & \cdots & 0\end{pmatrix}$$

令 $L_0=\begin{pmatrix}1 & & \\ l_{21} & 1 & \\ \vdots & \ddots & \\ & & 1 \\ \vdots & & \vdots \\ \vdots & & \vdots \\ l_{m1} & \cdots & l_{mr}\end{pmatrix}_{m\times r}$，$U_0=\begin{pmatrix}a_{11}^{(0)} & \cdots & \cdots & \cdots & a_{1n}^{(0)} \\ & \ddots & & & \vdots \\ & & a_{r,r}^{(r-1)} & \cdots & a_{r,n}^{(r-1)}\end{pmatrix}$，于是 $A=L_0U_0$

即为 A 的 LU 分解，(2)⇒(1)获证.

(1)⇒(2)　假定让 $A=LU$ 为 A 的 LU 分解，于是 $A=\left(L,\begin{pmatrix}0\\ I_{m-r}\end{pmatrix}\right)\begin{pmatrix}U\\ 0\end{pmatrix}$，从而 $\left(L,\begin{pmatrix}0\\ I_{m-r}\end{pmatrix}\right)^{-1}A=\begin{pmatrix}U\\ 0\end{pmatrix}$. 由于 U 的前 r 个顺序主子式均不为 0，且注意到 $\left(L,\begin{pmatrix}0\\ I_{m-r}\end{pmatrix}\right)^{-1}$ 左乘 A 实际上相当于对 A 的行施行了适当的第二类行初等变换，它不改变行列式值，因此 A 的前 r 个顺序主子式与 U 的前 r 个顺序主子式相同，(1)⇒(2)获证.

最后证明 LDU 分解的唯一性，从而也证明了 LU 分解的唯一性. 设：

$$L=\begin{pmatrix}1 & & \\ \vdots & \ddots & \\ l_{r1} & & 1 \\ \vdots & & \vdots \\ l_{m1} & \cdots & l_{mr}\end{pmatrix}_{m\times r},\ U=\begin{pmatrix}1 & u_{12} & \cdots & \cdots & u_{1n} \\ & \ddots & & & \\ & & 1 & \cdots & u_{rn}\end{pmatrix},\ D=\begin{pmatrix}d_1 & & \\ & \ddots & \\ & & d_r\end{pmatrix}$$

$A=LDU$ 为 A 的 LDU 分解. 通过具体计算 L、D、U 中的各元素，从而给出唯一性的证明.

首先，将 $A=LDU$ 进行如下分块：

$$\begin{pmatrix}A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix}=\begin{pmatrix}L_{11} & 0 \\ L_{21} & L_{22}\end{pmatrix}_{m\times r}\begin{pmatrix}D_1 & 0 \\ 0 & D_2\end{pmatrix}_{r\times r}\begin{pmatrix}U_{11} & U_{12} \\ 0 & U_{22}\end{pmatrix}_{r\times n}$$

其中 A_{11},L_{11},D_1,U_{11}分别是 A,L,D,U 前p行p列构成的子方阵,$p=1,2,\cdots,r$,于是

$$\left.\begin{array}{l}A_{11}=L_{11}D_1U_{11}\\A_{12}=L_{11}D_1U_{12}\end{array}\right\}\Rightarrow\begin{cases}A_{11}=L_{11}D_1U_{11}\\(A_{11},A_{12})=L_{11}D_1(U_{11},U_{12})\end{cases}$$

$$\left.\begin{array}{l}A_{21}=L_{21}D_1U_{11}\\A_{22}=L_{21}D_1U_{12}+L_{22}D_2U_{22}\end{array}\right\}\Rightarrow\begin{pmatrix}A_{11}\\A_{21}\end{pmatrix}=\begin{pmatrix}L_{11}\\L_{12}\end{pmatrix}D_1U_{11}$$

让 $A\begin{pmatrix}1,\cdots,p\\1,\cdots,p\end{pmatrix}$表示由$A$的第$1,\cdots$,第$p$行及第$1,\cdots$,第$p$列交叉处元素构成的方式.注意到:

$$A\begin{pmatrix}1,\cdots,p\\1,\cdots,p\end{pmatrix}=\det A_{11}=\det D_1=d_1\cdots d_p,\text{因此得}$$

$$d_1=A\begin{pmatrix}1\\1\end{pmatrix}=a_{11}$$

$$\vdots$$

$$d_p=\frac{A\begin{pmatrix}1,2,\cdots,p\\1,2,\cdots,p\end{pmatrix}}{A\begin{pmatrix}1,2,\cdots,p-1\\1,2,\cdots,p-1\end{pmatrix}}$$

$$p=1,2,\cdots,r$$

结果 D 被唯一确定.

考察(A_{11},A_{12})的前 $p-1$ 列和第 j 列$(j>p)$组成的子矩阵$(\tilde{A}_{11},a_j)$,则

$$\begin{cases}\det(\tilde{A}_{11},a_j)=A\begin{pmatrix}1,\cdots,p-1,p\\1,\cdots,p-1,j\end{pmatrix};\\(\tilde{A}_{11},a_j)=L_{11}D_1(\tilde{U}_{11},u_j)\end{cases}$$

其中 $\tilde{U}_{11},u_j$ 分别为(U_{11},U_{12})的前 $p-1$ 列和第 j 列.

结果

$$A\begin{pmatrix}1,\cdots,p-1,p\\1,\cdots,p-1,j\end{pmatrix}=\det D_1\det(\tilde{U}_{11},u_j)=d_1\cdots d_pu_{pj}$$

进而

$$u_{pj}=\frac{A\begin{pmatrix}1,\cdots,p-1,p\\1,\cdots,p-1,j\end{pmatrix}}{A\begin{pmatrix}1,2,\cdots,p\\1,2,\cdots,p\end{pmatrix}},\ j=p+1,2,\cdots,n$$

由此知

$$\begin{cases} 当\ p=1\ 时, u_{12},\cdots,u_{1n}被唯一确定, u_{1j}=\dfrac{A\begin{pmatrix}1\\j\end{pmatrix}}{A\begin{pmatrix}1\\1\end{pmatrix}} \\ \quad\vdots \\ 当\ p=r\ 时, u_{r,r+1},\cdots,u_{rn}被唯一确定, u_{rj}=\dfrac{A\begin{pmatrix}1,\cdots,r-1,r\\1,\cdots,r-1,j\end{pmatrix}}{A\begin{pmatrix}1,\cdots,r\\1,\cdots,r\end{pmatrix}},(j=r+1,\cdots,n) \end{cases}$$

U 是唯一的.

由于 U 有右逆, D 可逆,因此立即可知 L 也是唯一的.

□

由定理 2.4.1 立知:

推论 2.4.1 设 $A\in\mathbb{F}^{n\times n}$,则有以下两条等价:

(1) A 为 Hermite 正定阵;

(2) A 有唯一的分解 $A=LL^*$,其中 L 为 n 阶具有正对角元的下三角阵.

证明: (2)⇒(1) 显然.

(1)⇒(2) A 的正定性保证了 A 有唯一的 LDU 分解: $A=LDU$,其中 D 的对角元皆为正数.注意到 $A^*=A$,从而

$$U^*D^*L^*=U^*DL^*=LDU$$

结果 LDU 分解的唯一性导致 $L^*=U, A=LDL^*=(L\sqrt{D})(L\sqrt{D})^*$.

最后,若 $A=L_1L_1^*=L_2L_2^*$,其中 L_1,L_2 均为带正对角元的下三角阵,再次利用 LDU 分解的唯一性可知: $L_1=L_2$.

□

定义 2.4.2 设 $A\in\mathbb{F}^{n\times m}$, L 为带正对角元的下三角阵,若 $A=LL^*$,则称 $A=LL^*$ 为矩阵 A 的**Cholesky分解**.

显然矩阵的 Cholesky 分解实际上是针对 Hermite 正定阵而言的,并且它总是唯一的.

一般地,随意取定矩阵 $A\in\mathbb{F}^{n\times m}$,记 $r=\text{rank}A$,由定理 2.4.1 可知, A 的 LU 分解未必存在.为此,通过交换 A 的行和列,从而使 A 化为一个前 r 个顺序主子式皆非零的矩阵,再对之进行 LU 分解.在矩阵代数中,称此过程为矩阵 A 的选主元的 LU 分解.选主元的 LU 分解用矩阵语言加以描述即为:

$$PAQ=LU$$

其中 P,Q 皆为置换阵.

对一个 LU 分解存在的矩阵 A 来说,以下给出了矩阵的 LU 分解的初等行变换算法:

步骤 1 利用初等行变换将矩阵 A 化为行阶梯型矩阵 U;

步骤 2 对单位矩阵执行与步骤 1 相应的初等行逆变换,得到单位下三角阵;

输出 LU 分解由 $A=LU$ 依定义 2.4.1 给出.

关于上述算法,有两点值得指出:

第一点，步骤 2 的初等行逆变换是步骤 1 对应初等行变换的逆变换；第二点，若在步骤 1 中定义 $\alpha R_p + R_q$ 为将第 p 行元素乘常数 α 之后，加到第 q 行的初行变换，则步骤 2 中相应初等行逆变换 $(\alpha R_p + R_q)^{-1}$ 定义为 $-\alpha R_p + R_q$.

2.5　矩阵的 QR 分解

设 $A \in \mathbb{F}^{m\times n}(m \geqslant n)$，称分解 $A = QR$ 为 A 的一个 **QR 分解**是指该等式中，$Q \in \mathbb{F}^{n\times n}$ $(m \geqslant n)$ 为列正交的部分酉阵，而 $R \in \mathbb{F}^{m\times n}$ 则为上三角阵. 以下定理保证了任何矩阵的 QR 分解的存在性.

定理 2.5.1　设 $A \in \mathbb{F}^{m\times n}(m \geqslant n)$，则存在标准列正交的矩阵 $Q \in \mathbb{F}^{m\times n}$ 及上三角矩阵 $R \in \mathbb{F}^{n\times n}$，使得 $A = QR$. 如果 $m = n$，那么 Q 为酉阵；如果 A 为非奇异矩阵，则可以选取 R 为具有正对角元的上三角阵，并且在这种情况下，因子 Q 和 R 都是唯一的.

证明：(1) 记 $A = (a_{ij})_{m\times n} = (a_1, a_2, \cdots, a_n)$，由第 1 章的习题 2，可取适当的 Householder 变换阵 H_1，使得 $H_1 a_1 = r_{11} e_1$，$r_{11} \in \mathbb{F}$，e_1 为 m 维基本向量(即 $e_1 = (1, 0, \cdots, 1, 0)^{\mathrm{T}}$，$m$ 维)，此时

$$A_1 = H_1 A = [a_1^{(1)}, a_2^{(1)}, \cdots, a_n^{(1)}],\ a_1^{(1)} = r_{11} e_1;\ a_2^{(1)} = (a_{12}^{(1)}, a_{22}^{(1)}, \cdots, a_{m2}^{(1)})^{\mathrm{T}}$$

针对矩阵 A_1 的第 2 列 $a_2^{(1)}$，取适当的 Householder 变换阵 H_2，使得

$$H_2 x = r_{22} e_1$$

其中 $x = (a_{22}^{(1)}, \cdots, a_{m2}^{(1)})^{\mathrm{T}}$，$e_1$ 为 $m-1$ 维基本向量.

此时

$$A_2 \triangleq \begin{pmatrix} 1 & 0 \\ 0 & H_2 \end{pmatrix} A_1 = \begin{pmatrix} r_{11} & * & * & \cdots & * \\ 0 & r_{22} & * & \cdots & * \\ \vdots & 0 & * & \cdots & * \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & * & \cdots & * \end{pmatrix}$$

重复上述过程，则有

$$\begin{pmatrix} I & 0 \\ 0 & H_{n-1} \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ 0 & H_2 \end{pmatrix} H_1 A = \begin{pmatrix} r_{11} & * & \cdots & * \\ \vdots & & & \vdots \\ \vdots & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$$

令 $Q_1 = \begin{pmatrix} I & 0 \\ 0 & H_{n-1} \end{pmatrix} \cdots \begin{pmatrix} 1 & 0 \\ 0 & H_2 \end{pmatrix} H_1$，将 Q_1^{-1} 分块为 $Q_1^{-1} = (Q, Q_2)$，其中 Q 为 Q_1^{-1} 的前 n 列，将 $\begin{pmatrix} r_{11} & * & \cdots & * \\ \vdots & & & \vdots \\ \vdots & & & r_{nn} \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$ 分块为 $\begin{pmatrix} R \\ 0 \end{pmatrix}$，其中 R 为 $n\times n$ 阵，于是 $A = QR$.

(2)当 $m=n$ 时,(1)中的 Q 为酉阵是显然的.

(3)当 A 为可逆阵时,若 $A=QR$ 为 QR 分解,则 R 的对角元

$$r_{11}=\rho(r_{11})e^{i\theta_1},\cdots,r_{nn}=\rho(r_{nn})e^{i\theta_n}$$

均不为 0,于是

$$A=Q\begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_n} \end{pmatrix}\begin{pmatrix} \rho(r_{11}) & \cdots & * \\ & \ddots & \vdots \\ & & \rho(r_{nn}) \end{pmatrix}$$

此时 $Q\begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_n} \end{pmatrix}$仍为酉阵,$\begin{pmatrix} \rho(r_{11}) & \cdots & * \\ & \ddots & \vdots \\ & & \rho(r_{nn}) \end{pmatrix}$为具有正对角元的上三角阵.

最后,设 $A=Q_1R_1$ 与 $A=Q_2R_2$ 均为 A 的 QR 分解,R_1,R_2 均具有正对角元 Q_1,Q_2 均是酉阵.于是 $Q_2^*Q_1R_1=R_2$,$Q_2^*Q_1=R_2R_1^{-1}$ 为上三角阵,此时酉阵 $Q_2^*Q_1$ 变为对角阵,$Q_2^*Q_1$ 可表为:

$$Q_2^*Q_1=\begin{pmatrix} e^{i\beta_1} & & \\ & \ddots & \\ & & e^{i\beta_n} \end{pmatrix}$$

再次由 $Q_2^*Q_1R_1=\begin{pmatrix} e^{i\beta_1} & & \\ & \ddots & \\ & & e^{i\beta_n} \end{pmatrix}R_1=R_2$,结合 R_1,R_2 均具有正对角元,知 $e^{i\beta_1}=\cdots=e^{i\beta_n}=1$,结果 $Q_1=Q_2$,$R_1=R_2$.

□

在矩阵理论中,有时称下列的满秩分解为 QU 分解:

定义 2.5.1 设 $A\in\mathbb{F}^{m\times n}(m\geq n)$,$\mathrm{rank}A=r$. 若 $A=QU$,其中 Q 满足 $Q^*Q=I_r$,U 是秩为 r 的 $r\times n$ 上三角阵,则称 $A=QU$ 为 A 的一个 **QU 分解**,或称(Q,U)为 A 的一个 **QU 分解束**.

关于 QU 分解的存在性,易知下列定理成立.

定理 2.5.2 设 $A\in\mathbb{F}^{m\times n}(m\geq n)$,$\mathrm{rank}A=r$,且 A 的前 r 列线性无关,则 A 的 QU 分解存在.

证明: 类似定理 2.5.1 中(1)的证明即可.

□

与矩阵的 LU 分解不同,矩阵的 QU 分解未必唯一,由此也可看出,矩阵的 QR 分解一般也未必唯一.

定理 2.5.3 设 $A\in\mathbb{F}^{m\times n}(m\geq n)$,$\mathrm{rank}A=r$,且 A 的前 r 列线性无关. 若 $A=Q_1U_1$ 及 $A=Q_2U_2$ 均为 A 的 QU 分解,则存在对角阵

$$D=\begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_r} \end{pmatrix}$$

使得 $Q_1 = Q_2 D, U_1 = D^* U_2$,从而 $\{(Q_2 D, D^* U_2) \mid D = \begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_r} \end{pmatrix}, \theta_1, \cdots, \theta_r \in \mathbb{R}\}$ 便为 A 的所有 QU 分解束构成的集合.

证明: 由 $A = Q_1 U_1 = Q_2 U_2$ 知:

$$Q_2^* Q_1 U_1 = U_2, Q_2^* Q_1 \in \mathbb{F}^{r\times r}$$

$$Q_1 U_1 = Q_2 Q_2^* Q_1 U_1, Q_1 = Q_2 Q_2^* Q_1$$

结果令 $V = Q_2^* Q_1$,则有 $I_r = Q_1^* Q_1 = V^* Q_2^* Q_2 V = V^* V$, V 为酉阵.

将 U_1 分块为 $U_1 = (\underset{r\times r}{R_1}, S_1)$, U_2 分块为 $U_2 = (\underset{r\times r}{R_2}, S_2)$,于是

$$VU_1 = (VR_1, VS_1) = (R_2, S_2)$$

$VR_1 = R_2, V = R_2 R_1^{-1}$ 为上三角阵,进而 V 为对角矩阵, V 可表为

$$V = \begin{pmatrix} e^{i\theta_1} & & \\ & \ddots & \\ & & e^{i\theta_r} \end{pmatrix}$$

具体计算知: $Q_1 = Q_2 V, U_1 = V^* U_2$.

□

QR 分解有重要的数值意义,而且作为一种理论方法,它同样是有价值的,例如由 QR 分解,不难得出下列几个推论.

推论 2.5.1 设 B 为 Hermite 半正定 n 阶方阵,则 B 可以写成 $B = LL^*$,其中 L 是具有非负对角元的下三角阵.

证明: B 为 Hermite 半正定阵,说明 B 可以写成 $B = AA^*$ 的形式,于是取 A 的 QR 分解: $A = QR$,则 $B = R^* Q^* QR = R^* R$.

□

推论 2.5.2 设 $\mathscr{F}$ 为 $\mathbb{F}^{n\times n}$ 中的某个矩阵族,若 $\exists\, T \in \mathbb{F}^{n\times n}$,使得 $\forall A \in \mathscr{F}$,均有 TAT^{-1} 为上三角阵,则 $\mathscr{F}$ 必可经酉相似同时上三角化.

证明: 取 T^{-1} 的 QR 分解: $T^{-1} = QR$,于是 $TAT^{-1} = R^{-1} Q^{-1} AQR$ 为上三角阵,由上三角阵之积仍为上三角阵知: $Q^* AQ$ 为上三角阵.

□

习题二

1. 设 $A \in \mathbb{C}^{n\times n}$, $A = U\Sigma V^*$ 是 A 的奇异值分解,证明:

(1) 极大问题 $\max\{|\mathrm{tr}(AQ)| : Q \in \mathbb{C}^{n\times n}$ 是酉阵$\}$ 的值为 $\sum_{i=1}^{n} \sigma_i$,其中 σ_i 是矩阵A 的奇异值,而 $Q = VU^*$ 是问题的一个解.

(2)酉阵 Q 是上述极大问题解的充要条件是AQ 可表为 $AQ = Z \cdot H$,其中 H 为Hermite

正定阵,Z 为单位复数.

2. 设 $A\in\mathbb{C}^{m\times n}$,$B\in\mathbb{C}^{n\times m}$,求极大问题:

$$\max\{|\mathrm{tr}(AUBV)|:U\in\mathbb{C}^{n\times n},V\in\mathbb{C}^{m\times m}\text{是酉阵}\}.$$

3. 求下列矩阵的 LU 分解和 QR 分解.

(1) $\begin{pmatrix}1&2&3\\4&5&6\\2&4&6\end{pmatrix}$ (2) $\begin{pmatrix}1&3&2\\2&4&1\\4&6&0\end{pmatrix}$

4. 给定 n 阶复方阵 A,b 为 n 维列向量,非零多项式 $f(x)$ 若满足 $f(A)b=0$,则称 $f(x)$ 为**向量 b 关于 A 的化零多项式**,其中次数最低的称为**向量 b 关于 A 的最小多项式**. 证明:如果 φ_1,φ_2 分别是向量 b_1,b_2 关于矩阵 A 的化零多项式,则 $\varphi_1\varphi_2$ 就是 b_1,b_2 的任意线性组合 $b=kb_1+lb_2$ 的化零多项式.

5. 给定 n 阶复方阵 A,X 为 $\mathbb{C}^n$ 的非零线性子空间,非零多项式 $f(x)$ 若满足 $f(A)(X)=0$,则称 $f(x)$ 为**子空间 X 关于 A 的化零多项式**.在子空间 X 关于 A 的化零多项式中,次数最低的称为**子空间 X 关于 A 的最小多项式**.证明子空间 X 关于 A 的最小多项式必为 Z 中某向量关于 A 的最小多项式.

6. 给定 n 阶复方阵 A,b 为 n 维列向量,称 $\mathrm{span}\{b,Ab,\cdots,A^kb,\cdots\}=S_b$ 为 **b 关于 A 所产生的循环子空间**.证明:A 的非零不变线性子空间 X 必能分解为有限个 b 关于 A 所产生的循环子空间的直和.

7. 给定 n 阶复方阵 A,b 为 n 维列向量,$\mathrm{span}\{b,Ab,\cdots,A^kb,\cdots\}=S_b$ 关于 A 的最小多项式 $\varphi(\lambda)$ 具有形式 $\varphi(\lambda)=[\phi(\lambda)]^l$,其中 $\phi(\lambda)$ 是复数域上的不可约多项式,l 是正整数.证明:S_b 不可以再分解为 A 的两个非 0 不变子空间直和.

8. 设 $A\in\mathbb{C}^{n\times n}$,$\lambda$ 为 A 的一个特征值,r 为 λ 的指标.证明:

$$\dim N(A-\lambda I)\cap\mathrm{Im}(A-\lambda I)^{r-1}=\dim N(A-\lambda I)^r-\dim N(A-\lambda I)^{r-1}$$

9. 证明:$\mathbb{C}^n$ 的任意一个不等于 $\mathbb{C}^n$ 的非零子空间都是若干个 $n-1$ 维子空间的交.

10. 证明:一个 $2k$ 阶具有共轭特征值的实Jordan块与两个 k 阶特征值共轭的Jordan块构成的对角块阵是相似的.

第3章　矩阵的广义逆

每个可逆方阵 A 都有唯一的逆，常记作 A^{-1}. A^{-1}满足 $AA^{-1}=A^{-1}A=I$. 可逆矩阵又叫**非奇异矩阵**，它具有诸多好的性质和用途. 遗憾的是只有当 A 为方阵且为可逆矩阵时 A^{-1}才存在，而在应用数学的许多领域中，却常常遇到长方阵和奇异方阵，并且为满足实际需要，对于这样一类矩阵，又希望能构造出具有通常逆矩阵的若干性质的矩阵，这样的矩阵在矩阵代数中称为**广义逆**，有时也称为**伪逆**. 沿着这一思路，数学家 E. H. Moore 于 1929 年美国数学会上提出了他的广义逆矩阵的一个论文摘要，其论文发表在他死后的 1935 年. 20 世纪 30 年代，我国的曾远荣先生还将之推广到 Hilbert 空间线性算子中. 20 世纪 50 年代，由于数学的发展，广义逆理论开始受到重视. 1955 年，R. Penrose 发表了与 Moore 等价的广义逆理论，现称为矩阵的 **Moore - Penrose 逆**，也叫**加号逆**，常记作 A^{+}. 同年，Rao 提出了一个更一般广义逆矩阵的概念，现在叫做 **g 逆**，常记作 A^{-} 或 $A^{\{1\}}$，本书中也称之为**{1} - 逆**. 目前，广义逆理论已广泛应用于诸如数据处理、多元分析、最优化理论、现代控制理论、网络理论等各个学科中.

3.1　广义逆矩阵 A^{-} 及其一般表达式

定义 3.1.1　设 $A\in\mathbb{F}^{m\times n}$, $G\in\mathbb{F}^{n\times m}$，若 G 满足 $AGA=A$，则称 G 为 A 的一个广义逆，也称为**{1} - 逆**，常记作 A^{-} 或 $A^{\{1\}}$. A 的所有 {1} - 逆的集合记为 $A\{1\}$；若 G 满足 $GAG=G$，则称 G 为 A 的一个**{2} - 逆**，常记作 $A^{\{2\}}$. A 的所有 {2} - 逆的集合记为 $A\{2\}$；若 G 既是 A 的 {1} - 逆又是 A 的 {2} - 逆，则记为 $G\in A\{1,2\}$.

定理 3.1.1　设 $A\in\mathbb{F}^{m\times n}$，则 A 的 {1} - 逆必存在，A 的 {2} - 逆也必存在，且 $A\{1,2\}\neq\emptyset$.

证明: 令 $A=P\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}Q$，其中 P,Q 可逆，于是有 $G=Q^{-1}\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}_{n\times m}P^{-1}$，满足 $AGA=A$. 现取一个 $G\in A\{1\}$，则 $AGA=A$. 让 $G_0=GAG$，则 $G_0AG_0=G_0$, $G_0\in A\{2\}\neq\emptyset$. 进一步对上述 G_0 而言，$AG_0A=AGAGA=A$，因此 $A\{1,2\}\neq\emptyset$.

□

例 3.1.1　设 $A\in\mathbb{F}^{n\times n}$，$A$ 可逆，则 $A\{1\}=\{A^{-1}\}$, $A\{1,2\}=\{A^{-1}\}$.

定理 3.1.2　设 $A\in\mathbb{F}^{m\times n}$. A^{-} 为 A 的一个 {1} - 逆，则有

$$A\{1\}=\{A^{-}+X(I_m-AA^{-})+(I_n-A^{-}A)Y\mid X,Y\in\mathbb{F}^{n\times m}\}$$

证明：(1) $\forall X,Y\in\mathbb{F}^{n\times m}$，有

$$
\begin{aligned}
&A(A^- + X(I_m - AA^-) + (I_n - A^-A)Y)A\\
&= AA^-A + AX(I_m - AA^-)A + A(I_n - A^-A)YA\\
&= AA^-A + 0 + 0\\
&= A.
\end{aligned}
$$

(2)设 $G\in A\{1\}$，令 $X = G - A^-$，$Y = GAA^-$，则

$$
\begin{aligned}
&A^- + X(I_m - AA^-) + (I_n - A^-A)Y\\
&= A^- + (G - A^-)(I_m - AA^-) + (I_n - A^-A)GAA^-\\
&= G
\end{aligned}
$$

□

对 n 阶方阵 $A\in\mathbb{F}^{n\times n}$，若 $\exists\, G\in\mathbb{F}^{n\times n}$ 使得 $AG = I_n$，则称 A 可逆，G 为 A 的逆矩阵 A^{-1}．当 A 可逆时，其逆矩阵存在唯一．对 $A\in\mathbb{F}^{m\times n}$，若 $G\in\mathbb{F}^{n\times m}$ 满足 $AG = I_m$，则称 G 为 A 的一个**右逆**，记为 A_R^{-1}．类似地有 A 的**左逆**概念．A 的左逆记为 A_L^{-1}．一般地，$\forall A\in\mathbb{F}^{m\times n}$，其 A_R^{-1}，A_L^{-1} 未必存在．

定理 3.1.3 设 $A\in\mathbb{F}^{m\times n}$，有

(1) A_R^{-1} 存在 $\Leftrightarrow$ 秩 $A = m$（行满秩），此时 A_R^{-1} 可取为 $A^*(AA^*)^{-1}$；

(2) A_L^{-1} 存在 $\Leftrightarrow$ 秩 $A = n$（列满秩），此时 A_L^{-1} 可取为 $(A^*A)^{-1}A^*$；

(3) $A_R^{-1} = A_L^{-1} \Leftrightarrow A$ 可逆，此时，$A^{-1} = A_R^{-1} = A_L^{-1}$．

证明：(1)设 A_R^{-1} 存在，则 $AA_R^{-1} = I_m$，将 A 写成 $A = \begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_m\end{pmatrix}$．

若 $k_1\alpha_1 + \cdots + k_m\alpha_m = 0$，则 $(k_1,\cdots,k_m)\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_m\end{pmatrix}A_R^{-1} = 0$，于是 $(k_1,\cdots,k_m)I_m = 0$．亦即有 $k_1 = \cdots = k_m = 0$．故秩 $A = m$，行满秩．

反之，当秩 $A = m$ 时，A^*A 为 m 阶方阵，且秩(A^*A) = 秩 $A = m$．故

$$AA^*(AA^*)^{-1} = I_m,\ A^*(AA^*)^{-1}$$

为 A 的右逆．

(2)类似可证．

(3)由方阵可逆的定义即得．

□

矩阵的左、右逆具有下列性质：

定理 3.1.4 设 $A\in\mathbb{F}^{m\times n}$，$A_R^{-1}$ 表示 A 的一个右逆，A_L^{-1} 表示 A 的一个左逆，则以下几条成立：

(1) $AA_R^{-1}A = A$，$AA_L^{-1}A = A$；

(2) $A_R^{-1}AA_R^{-1} = A_R^{-1}$，$A_L^{-1}AA_L^{-1} = A_L^{-1}$；

(3) $(AA_R^{-1})^* = AA_R^{-1}, (A_L^{-1}A)^* = A_L^{-1}A$.

证明: 逐条验证即可.

□

设 $A \in \mathbb{F}^{m\times n}$, 关于 A 的广义逆的性质可列出如下几条:

性质 3.1.1　$(A^-)^* \in A^*\{1\}$.

证明: $AA^-A = A \Leftrightarrow A^*(A^-)^*A^* = A^*$, 因此 $(A^-)^* \in A^*\{1\}$.

□

性质 3.1.2　$A(A^*A)^-(A^*A) = A; A^*A(A^*A)^-A^* = A^*$.

证明: 注意到对任何矩阵 X, 若 $X^*X = 0$, 则 $X = 0$, 这归功于秩 $X^*X =$ 秩 X. 因此

$$\begin{aligned}&[A(A^*A)^-(A^*A) - A]^*[A(A^*A)^-(A^*A) - A]\\ =&[(A^*A)(A^*A)^-A^* - A^*][A(A^*A)^-(A^*A) - A]\\ =&A^*A - A^*A - A^*A + A^*A\\ =&0\end{aligned}$$

注意:这里第一个等号后面的两个 $(A^*A)^-$ 未必相同.

因此 $A(A^*A)^-(A^*A) = A^*$.

类似地可证 $A^*A(A^*A)^-A^* = A^*$.

□

性质 3.1.3　$AGA = A \Leftrightarrow A^*AGA = A^*A$.

证明: 由第 1 章的习题 3 立即可得.

□

性质 3.1.4　$\text{rank}A^- \geqslant \text{rank}A$

证明: 由 $AA^-A = A$ 立即得.

□

前面的定理 3.1.2 给出了一个矩阵的广义逆的一般表达式,但它是基于在事先已给出了一个 A^- 的情形下.下面的结果则克服了这一不足.

定理 3.1.5　设 $A \in \mathbb{F}^{m\times n}, A = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q, P, Q$ 可逆,则

$$A\{1\} = \left\{ Q^{-1}\begin{pmatrix} I_r & C \\ D & F \end{pmatrix}P^{-1} \mid C \in \mathbb{F}^{r\times(m-r)}, D \in \mathbb{F}^{(n-r)\times r}, F \in \mathbb{F}^{(n-r)\times(m-r)} \right\}$$

证明: 首先
$$\begin{aligned}AQ^{-1}\begin{pmatrix} I_r & C \\ D & F \end{pmatrix}P^{-1}A &= P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}QQ^{-1}\begin{pmatrix} I_r & C \\ D & F \end{pmatrix}P^{-1}P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q\\ &= P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} I_r & C \\ D & F \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q\\ &= P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q = A\end{aligned}$$

其次,若 $G \in A\{1\}$, 令 $QGP = \underset{\quad r \qquad m-r}{\begin{pmatrix} B & C \\ D & F \end{pmatrix}}\begin{matrix} r \\ n-r \end{matrix}$, 则

$$P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q = AGA = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} QGP\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q$$

$$= P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} B & C \\ D & F \end{pmatrix}\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q$$

结果　$B = I_r$.

□

回顾对一个 $A\in \mathbb{F}^{m\times n}$, $A = GH$ 叫做 A 的一个**满秩分解**是指：G 为列满秩阵而 H 为行满秩阵.

命题 3.1.1　设 $A\in \mathbb{F}_r^{m\times n}$, $A = BC$ 和 $A = \tilde{B}\tilde{C}$ 均为矩阵 A 的满秩分解. 记 $S_1 = \{C_R^{-1}B_L^{-1} \mid C_R^{-1}$为 C 的右逆, B_L^{-1} 为 B 的左逆$\}$, $S_2 = \{(\tilde{C})_R^{-1}(\tilde{B})_L^{-1} \mid (\tilde{C})_R^{-1}$ 为 $\tilde{C}$ 的右逆, $(\tilde{B})_L^{-1}$ 为 $\tilde{B}$ 的左逆$\}$, 则有 $S_1 = S_2$.

为证明此命题,先给出如下引理.

引理 3.1.1　设 $A\in \mathbb{F}_r^{m\times n}$, $A = BC$ 为 A 的一个满秩分解, 则$\{(BQ, Q^{-1}C) \mid Q$为可逆 r 阶方阵$\}$为由 A 的所有满秩分解对构成的集合.

证明： 只需证若 $A = BC = \tilde{B}\tilde{C}$ 均为 A 的满秩分解, 则存在可逆阵 Q 使得

$$B = \tilde{B}Q, C = Q^{-1}\tilde{C}$$

即可. 为此, 由 $BC = \tilde{B}\tilde{C}$ 得

$$BCC^* = \tilde{B}\tilde{C}C^*$$

$$B = \tilde{B}\tilde{C}C^* \cdot (CC^*)^{-1}$$

令 $Q_1 = \tilde{C}C^*(CC^*)^{-1}$, 则 $Q_1\in \mathbb{F}^{r\times r}$, $\tilde{B}Q_1 = B$. 下面验证 Q_1^{-1} 存在且 $Q_1^{-1}\tilde{C} = C$.

由 $B^*BC = B^*\tilde{B}\tilde{C}$ 得 $C = (B^*B)^{-1}B^*\tilde{B}\tilde{C}$, 从而令 $Q_2 = (B^*B)^{-1}B^*\tilde{B}$ 后有 $C = Q_2\tilde{C}$, 于是 $BC = \tilde{B}Q_1Q_2\tilde{C} = \tilde{B}\tilde{C}$, 这导致 $Q_1Q_2 = I_r$, 这说明 Q_1 可逆, $Q_2 = Q_1^{-1}$, 引理获证.

□

利用上述引理来证明命题 3.1.1.

命题 3.1.1 的证明　由引理 3.1.1, 存在可逆阵 Q, 使得 $B = \tilde{B}Q$, $C = Q^{-1}\tilde{C}$. 于是, $\forall X\in S_1$, $X = C_R^{-1}B_L^{-1}$, 从而

$$(QB_L^{-1})\tilde{B} = QB_L^{-1}\cdot BQ^{-1} = QQ^{-1} = I_r$$

$$\tilde{C}(C_R^{-1}Q^{-1}) = (QC)(C_R^{-1}Q^{-1}) = QQ^{-1} = I_r$$

因此 QB_L^{-1} 为 $\tilde{B}$ 的左逆, $C_R^{-1}Q^{-1}$为 $\tilde{C}$ 的右逆, 进而 $X = C_R^{-1}B_L^{-1}$ 为 S_2 中的元. 对称地有 $S_2 \subseteq S_1$. 命题获证.

□

现给定 $A\in \mathbb{F}^{m\times n}$, 能轻松地找到可逆 $R\in \mathbb{F}^{m\times n}$和可逆 $Q\in \mathbb{F}^{n\times n}$, 使得

$$A = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q$$

因此, 总有

$$A = P\begin{pmatrix} I_r \\ 0 \end{pmatrix}(I_r, 0)Q$$

为 A 的一个满秩分解 GH,其中 $G=P\begin{pmatrix}I_r\\0\end{pmatrix}$, $H=(I_r,0)Q$.现在可给出求已知矩阵的一个广义逆的方法.

方法一　将 A 进行满秩分解 $A=GH$,则任取 G 的一个左逆 G_L^{-1},任取 H 的一个右逆 H_R^{-1},于是 $H_R^{-1}G_L^{-1}$ 满足 $AH_R^{-1}G_L^{-1}A=A$,即 $H_R^{-1}G_L^{-1}$ 为 A 的一个广义逆.特别地,对列满秩阵 G 可取 $G_L^{-1}=(G^*G)^{-1}G^*$,对行满秩阵 H 可取 $H_R^{-1}=H^*(HH^*)^{-1}$.

方法二　对矩阵 A 施行一系列行初等变换,即左乘一个可逆矩阵 P,使得 $PA=\begin{pmatrix}C\\0\end{pmatrix}$,记 $A_1=\begin{pmatrix}C\\0\end{pmatrix}$,其中 C 为 $r\times n$ 阶矩阵,$\operatorname{rank}C=r$,则 A_1^-P 为 A 的一个广义逆,其中 $C_R^{-1}=C^*(CC^*)^{-1}$, $A_1^-=(C_R^{-1},0)_{n\times m}$.

证明: (1) $\begin{pmatrix}C\\0\end{pmatrix}(C_R^{-1},0)\begin{pmatrix}C\\0\end{pmatrix}=\begin{pmatrix}I_r&0\\0&0\end{pmatrix}\begin{pmatrix}C\\0\end{pmatrix}=\begin{pmatrix}C\\0\end{pmatrix}=A_1$;

(2) $AA_1^-PA=P^{-1}\begin{pmatrix}C\\0\end{pmatrix}(C_R^{-1},0)PP^{-1}\begin{pmatrix}C\\0\end{pmatrix}=P^{-1}A_1=A$.

□

方法三　对 A 施行一系列列的初等变换,即右乘一个可逆阵 Q 使得

$$AQ=(C,0),\text{记}(C,0)=A_1$$

其中 C 是 $m\times r$ 阶列满秩阵,则 QA_1^- 为 A 的一个广义逆,其中 $A_1^-=\begin{pmatrix}C_L^{-1}\\0\end{pmatrix}$, $C_L^{-1}=(C^*C)^{-1}C^*$.

例 3.1.2　设 $A=\begin{pmatrix}1&0&2\\0&1&0\\1&0&2\\1&0&2\end{pmatrix}_{4\times 3}$,求 A 的一个广义逆.

解:方法一　将 A 通过初等变换得出其一个满秩分解:

$$(A,I_4)=\begin{pmatrix}1&0&2&1&0&0&0\\0&1&0&0&1&0&0\\1&0&2&0&0&1&0\\1&0&2&0&0&0&1\end{pmatrix}\to\begin{pmatrix}1&0&2&1&0&0&0\\0&1&0&0&1&0&0\\0&0&0&-1&0&1&0\\0&0&0&-1&0&0&1\end{pmatrix}$$

于是 $PA=\begin{pmatrix}1&0&2\\0&1&0\\0&0&0\\0&0&0\end{pmatrix}$,其中 $P=\begin{pmatrix}1&0&0&0\\0&1&0&0\\-1&0&1&0\\-1&0&0&1\end{pmatrix}$,将 $\begin{pmatrix}1&0&2\\0&1&0\\0&0&0\\0&0&0\end{pmatrix}$ 进行列初等变换:

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

于是$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$,其中 $Q = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.现在 $PAQ = \begin{pmatrix} I_2 & 0 \\ 0 & 0 \end{pmatrix}$,从而得出 A 的一个满秩分解:

$$A = P^{-1}\begin{pmatrix} I_2 \\ 0 \end{pmatrix}(I_2 \quad 0)Q^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix} = GH$$

其中 $G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}$,$H = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}$.取 $H_R^{-1} = H^*(HH^*)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix}\begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \\ \frac{2}{5} & 0 \end{pmatrix}$.

再取

$$G_L^{-1} = (G^*G)^{-1}G^* = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

于是得到 A 的一个$\{1\}$-逆为 $H_R^{-1}G_L^{-1} = \begin{pmatrix} \frac{1}{15} & 0 & \frac{1}{15} & \frac{1}{15} \\ 0 & 1 & 0 & 0 \\ \frac{2}{15} & 0 & \frac{2}{15} & \frac{2}{15} \end{pmatrix}$.

方法二 将 A 通过行初等变换,将其化为阶梯形:

$$(A \quad I_4) = \begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}$$

于是$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} C \\ 0 \end{pmatrix}$,其中 $C = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}$.

记

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}$$

$$A_1 = \begin{pmatrix} C \\ 0 \end{pmatrix}$$

$$A_1^- = (C_R^{-1}, 0)_{3\times 4}$$

$$C_R^{-1} = C^*(CC^*)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} (CC^*)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} \left[\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} \right]^{-1}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \\ \frac{2}{5} & 0 \end{pmatrix}$$

因此 $A_1^- = \begin{pmatrix} \frac{1}{5} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{2}{5} & 0 & 0 & 0 \end{pmatrix}$,从而

$$A_1^- P = \begin{pmatrix} \frac{1}{5} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{2}{5} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{2}{5} & 0 & 0 & 0 \end{pmatrix}$$

便是 A 的一个广义逆.

□

3.2 Moore - Penrose 逆

在对非奇异阵的逆矩阵的推广过程中,Moore - Penrose 逆最具有代表性.

定义 3.2.1 给定 $A \in \mathbb{F}^{m\times n}$,满足下列四个方程的矩阵 $G \in \mathbb{F}^{n\times m}$,叫做 A 的 Moore - Penrose 逆,也叫做加号逆,记作 A^+:

(1) $AGA=A$;

(2) $GAG=G$;

(3) $(AG)^*=AG$;

(4) $(GA)^*=GA$.

同定义 A 的$\{1,2\}$ - 逆类似,满足(1)、(3)方程的矩阵 G 叫做 A 的**{1,3} - 逆**.满足方程(1)、(4)的矩阵 G 叫做 A 的**{1,4} - 逆**.通常 A 的$\{1,2\}$ - 逆又叫做 A 的**反射逆**;A 的$\{1,3\}$ - 逆又叫做 A 的**最小二乘 g 逆**,常记作 A_l^-;A 的$\{1,4\}$ - 逆又叫做 A 的**极小范数 g 逆**,常记作 A_m^-;A 的加号逆又叫做 A 的**极小最小二乘 g 逆**.

定理 3.2.1 对任意的 $A\in\mathbb{F}^{m\times n}$,$A^+$ 存在且唯一.

证明:(1)给定 $A\in\mathbb{F}^{m\times n}$,先证明 A^+ 存在.

取 A 的一个满秩分解 $A=CD$,再取 D_R^{-1} 为 $D_R^{-1}=D^*(DD^*)^{-1}$,C_L^{-1} 为 $C_L^{-1}=(C^*C)^{-1}C^*$,于是考察 $D_R^{-1}C_L^{-1}$,有

$$\begin{aligned}&([D^*(DD^*)^{-1}(C^*C)^{-1}C^*]A)^*\\=&[D^*(DD^*)^{-1}(C^*C)^{-1}C^*CD]^*\\=&[D^*(DD^*)^{-1}D]^*\\=&D^*(DD^*)^{-1}D\\=&D_R^{-1}C_L^{-1}A\end{aligned}$$

因此 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*\in A\{1,4\}$.

进一步验算知:

$D^*(DD^*)^{-1}(C^*C)^{-1}C^*\in A\{1,3\}$,且 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*\in A\{2\}$.所以 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*$ 为 A 的一个加号逆.

(2) A^+ 的唯一性.

设 G_1,G_2 均满足方程(1)、(2)、(3)、(4),则

$$\begin{aligned}G_1&=G_1AG_1=G_1AG_2AG_1=(G_1A)(G_2A)G_1=(G_1A)^*(G_2A)^*G_1\\&=A^*G_1^*A^*G_2^*G_1=(AG_1A)^*G_2^*G_1=A^*G_2^*G_1=(G_2A)^*G_1\\&=G_2AG_1=G_2AG_2AG_1=G_2(AG_2)^*(AG_1)^*=G_2G_2^*A^*G_1^*A^*\\&=G_2G_2^*(AG_1A)^*=G_2G_2^*A^*=G_2(AG_2)^*=G_2AG_2=G_2\end{aligned}$$

□

定理 3.2.2 给定 $A\in\mathbb{F}^{m\times n}$,则

(1) $A\{1,3\}=\{A_l^-+(I_n-A_l^-A)Z\mid Z\in\mathbb{F}^{n\times m}$,$A_l^-$ 为任意事先给定的一个 $A\{1,3\}$ - 逆$\}$.例如,通过取 A 的一个满秩分解 $A=CD$,那么 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*$ 便是一个确定的 $A\{1,3\}$ - 逆.

(2) $A\{1,4\}=\{A_m^-+Z(I_m-AA_m^-)\mid Z\in\mathbb{F}^{n\times m}$,$A_m^-$ 为任意事先给定的一个 $A\{1,4\}$ - 逆$\}$.例如,取 A_m^- 为(1)中的 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*$.

证明:(1) $\forall Z\in\mathbb{F}^{n\times m}$,有

$$A[A_l^-+(I_n-A_l^-A)Z]A=AA_l^-A+A(I_n-A_l^-A)ZA=A+0=A$$

$$(A[A_l^-+(I_n-A_l^-A)Z])^*=[AA_l^-+0]^*=AA_l^-=A[A_l^-+(I_n-A_l^-A)Z]$$

另外,若 $G\in A\{1,3\}$,则有

$$AA_l^- = AGAA_l^- = (AG)^*(AA_l^-)^* = G^*A^*(A_l^-)^*A^*$$
$$= G^*(AA_l^-A)^* = G^*A^* = (AG)^* = AG$$

令 $Z = G - A_l^-$,于是

$$A_l^- + (I_n - A_l^-A)Z = A_l^- + (I_n - A_l^-A)(G - A_l^-)$$
$$= A_l^- + G - A_l^-AG - A_l^- + A_l^-AA_l^- = G - A_l^-AA_l^- + A_l^-AA_l^-$$
$$= G$$

(2)显然有 $\forall Z\in \mathbb{F}^{n\times m}, A_m^- + Z(I_m - AA_l^-)\in A\{1,4\}$.另外, $\forall G\in A\{1,4\}$,则有

$$A_m^-A = A_m^-AGA = (A_m^-A)^*(GA)^* = A^*(A_m^-)^*A^*G^*$$
$$= (AA_m^-A)^*G^* = (GA)^* = GA$$

因此令 $Z = G - A_m^-$,于是

$$A_m^- + Z(I_m - AA_m^-) = A_m^- + (G - A_m^-)(I_m - AA_m^-)$$
$$= A_m^- + G - A_m^- - GAA_m^- + A_m^-AA_m^-$$
$$= G - GAA_m^- + A_m^-AA_m^-$$
$$= G - A_m^-AA_m^- + A_m^-AA_m^-$$
$$= G$$

□

从定理3.2.2的证明不难看出,给定 $A\in \mathbb{F}^{m\times n}$, A_m^-, A_l^- 分别为事先给定的任一 $A\{1,4\}$ - 逆, $A\{1,3\}$ - 逆,则有

$$G\in A\{1,4\} \Leftrightarrow GA = A_m^-A$$

$$G\in A\{1,3\} \Leftrightarrow AG = AA_l^-$$

综合上述定理及本章的定理3.1.2便知,给定矩阵 $A\in \mathbb{F}^{m\times n}$,则 $A\{1\}, A\{1,3\}, A\{1,4\}$的通式均能以 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*$ 为起点,其中 C, D 满足 $A = CD$ 是 A 的任一满秩分解.进一步,矩阵 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*$ 还为 A 的唯一的 Moore - Penrose 逆.事实上,基于上述的 $D^*(DD^*)^{-1}(C^*C)^{-1}C^*$,还可以写出 $A\{1,2\}$的通式如下:

$$A\{1,2\} = \{G_1AG_2 \mid G_1, G_2\in A\{1\}\}$$

证明: $G_1AG_2\in A\{1,2\}$是显然的.另一方面,若 $G\in A\{1,2\}$,则有$\begin{cases} AGA = A \\ GAG = G \end{cases}$.于是由等式 $GAG = G$ 直接知: $G\in \{G_1AG_1 \mid G_1, G_2\in A\{1\}\}$.

□

一般地,用满秩分解来分别求出一个矩阵的一个$\{1,4\}$ - 逆、$\{1,3\}$ - 逆时,尽管理论上可行,但实际操作起来却较为麻烦.通常可使用如下方法来给出一个矩阵的一个$\{1,4\}$逆和一个$\{1,3\}$ - 逆,甚至 A^+:

$\forall A\in \mathbb{F}^{m\times n}$,有 $A_m^- = A^*(AA^*)^-$; $A_l^- = (A^*A)^-A^*$; $A^+ = A_m^-AA_l^-$,其中$(AA^*)^-$和$(A^*A)^-$则通过本章的3.1节中的方法二或方法三来得到.这完全归功于下述命题成立.

命题 3.2.1 设 $A\in\mathbb{F}^{m\times n}$,则

(1)对 AA^* 的任两个$\{1\}$-逆 X 和 Y,总有 $A^*XA=A^*YA$;

(2)对 A^*A 的任两个$\{1\}$-逆 X 和 Y,也总有 $AXA^*=AYA^*$.

证明: 只证(2). $Y\in A^*A\{1\}$,由 3.1 节中的性质 3.1.2 得

$$\begin{cases}(A^*A)YA^*=A^*\\ AYA^*A=A\end{cases}$$

$$AXA^*=(AYA^*A)XA^*=AY(A^*AXA^*)=AYA^*$$

□

例 3.2.1 设 $A=\begin{pmatrix}3&0&5&6\\0&0&7&1\\1&0&0&0\end{pmatrix}$,求出 A^+.

解: (1) $AA^*=\begin{pmatrix}70&41&3\\41&50&0\\3&0&1\end{pmatrix}$

(2) AA^* 的$\{1\}$-逆只有 1 个,它为 $(AA^*)^{-1}=\begin{pmatrix}\frac{50}{1369}&\frac{8255}{1369}&\frac{5526}{1369}\\ \frac{-41}{1369}&\frac{-61}{1369}&\frac{-41}{1369}\\ \frac{16578}{1369}&\frac{123}{1369}&\frac{1519}{1369}\end{pmatrix}$

(3) $A^+=A^*(AA^*)^{-1}=\begin{pmatrix}\frac{16728}{1369}&\frac{24888}{1369}&\frac{18097}{1369}\\ 0&0&0\\ \frac{-37}{1369}&\frac{40848}{1369}&\frac{27343}{1369}\\ \frac{259}{1369}&\frac{49469}{1369}&\frac{33115}{1369}\end{pmatrix}$

注意:本例用到下述事实:若 A 为行满秩阵,则必有 $A^+=A^*(AA^*)^{-1}$.关于这一事实的证明,我们留作习题.

3.3 矩阵广义逆在求解线性方程组中的应用

设 $A\in\mathbb{F}^{m\times n}$,所谓相容线性方程组是指有解的线性方程组 $AX=b$,而 $AX=b$ 有解的充要条件又是秩 A = 秩(A,b).矩阵 A 的$\{1\}$-逆能较好地刻画相容方程组 $AX=b$ 的解情况.

定理 3.3.1(相容方程组的通解) 相容线性方程组 $AX=b$ 的通解是

$$X=A^-b+(I_n-A^-A)Y$$

其中 A^- 为 A 的任意取定的一个$\{1\}$-逆,Y 为$\mathbb{F}^n$ 中的任意元素,是自由变量.

证明：

(1)$A[A^- b+(I_n-A^- A)Y]=AA^- b+0$.因为 $AX=b$ 相容，b 可以写成 AX_0 的形式，所以 $AA^- b=AA^- AX_0=AX_0=b$.

(2)设 X_0 为 $AX=b$ 的任何一个解，取 $Y=X_0$，于是

$$A^- b+(I_n-A^- A)X_0=A^- b+X_0-A^- AX_0=X_0$$

□

定理 3.3.2 (1)相容线性方程组 $AX=b(b\neq 0)$的解集为$\{Gb\mid G\in A\{1\}\}$；

(2)线性方程组 $AX=0$ 的解集为$\{(I_n-A^- A)Y\mid Y\in \mathbb{F}^n\}$，其中 A^- 为 A 的任意取定的一个$\{1\}$-逆.

证明：(1)设 X_0 为 $AX=b(b\neq 0)$的任一解.让 $A=P\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}Q$，其中 P 为 m 阶可逆阵，Q 为 n 阶可逆方阵.将 QX_0 分块为$\begin{pmatrix} X_1\\ X_2\end{pmatrix}\begin{matrix} r\\ n-r\end{matrix}$，$P^{-1}b$ 分块为$\begin{pmatrix} Y_1\\ Y_2\end{pmatrix}\begin{matrix} r\\ m-r\end{matrix}$.于是 X_0 是 $AX=b$ 的一个解

$$\Leftrightarrow P\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}QX_0=b\Leftrightarrow \begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}QX_0=P^{-1}b$$

$$\Leftrightarrow \begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix} X_1\\ X_2\end{pmatrix}\begin{matrix} r\\ n-r\end{matrix}=\begin{pmatrix} Y_1\\ Y_2\end{pmatrix}\begin{matrix} r\\ m-r\end{matrix}\Leftrightarrow\begin{cases} X_1=Y_1\\ Y_2=0\end{cases}$$

因 $b\neq 0$，则 $P^{-1}b\neq 0$，从而 $Y_1\neq 0$，进而$\{DY_1\mid D\in\mathbb{F}^{(n-r)\times r}\}=\mathbb{F}^{(n-r)\times 1}$.因此可取 D_0 使得 $D_0Y_1=X_2$，结果：

$$Q^{-1}\begin{pmatrix} I_r & 0\\ D_0 & 0\end{pmatrix}P^{-1}b=Q^{-1}\begin{pmatrix} I_r & 0\\ D_0 & 0\end{pmatrix}\begin{pmatrix} Y_1\\ Y_2\end{pmatrix}=Q^{-1}\begin{pmatrix} Y_1\\ X_2\end{pmatrix}=Q^{-1}\begin{pmatrix} X_1\\ X_2\end{pmatrix}=Q^{-1}QX_0=X_0$$

令 $G=Q^{-1}\begin{pmatrix} I_r & 0\\ D_0 & 0\end{pmatrix}P^{-1}$，则由本章的定理 3.1.5 可知，$G\in A\{1\}$，$X_0=Gb$.结果 $AX=b(b\neq 0)$的解集为$\{Gb\mid G\in A\{1\}\}$.

(2)$AX=0$ 的解集为$\{(I_n-A^- A)Y\mid Y\in\mathbb{F}^n\}$是定理 3.3.1 的直接推论.

□

定理 3.3.3(Penrose 定理) 矩阵方程 $AXB=C$ 有解 $\Leftrightarrow$ $AA^- CB^- B=C$，并且在有解的情况下，其通解为 $X=A^- CB^- +Y-A^- AYBB^-$，其中 Y 为任意维数适当的矩阵.

证明：直接验证即可.

□

例 3.3.1 求齐次线性方程组$\begin{cases} x_1+2x_2-x_3=0\\ -x_2+2x_3=0\end{cases}$的通解.

解：将其写成矩阵形式有

$$AX=\begin{pmatrix} 1 & 2 & -1\\ 0 & -1 & 2\end{pmatrix}\begin{pmatrix} x_1\\ x_2\\ x_3\end{pmatrix}=\begin{pmatrix} 0\\ 0\\ 0\end{pmatrix}$$

A 的一个$\{1\}$-逆为$\frac{1}{14}\begin{pmatrix}5 & 4\\6 & 2\\3 & 8\end{pmatrix}$,所求线性方程组的通解为:

$$X=(I-A^{-}A)Y=\frac{1}{14}\begin{pmatrix}9c_1-6c_2-3c_3\\-6c_1+4c_2+2c_3\\-3c_1+2c_2+\ \ c_3\end{pmatrix}$$

其中 $Y=\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}$为$\mathbb{F}^3$ 中的任意元素.

□

例 3.3.2 求线性方程组$\begin{cases}x_1+2x_2-x_3=-2\\-x_2+2x_3=4\end{cases}$的通解.

解: 由上面的例 3.3.1 知,$\begin{pmatrix}1 & 2 & -1\\0 & -1 & 2\end{pmatrix}$的一个$\{1\}$-逆为$\frac{1}{14}\begin{pmatrix}5 & 4\\6 & 2\\3 & 8\end{pmatrix}$,于是由 Penrose 定理或定理 3.3.1 知所求方程的通解为

$$X=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\frac{1}{14}\begin{pmatrix}5 & 4\\6 & 2\\3 & 8\end{pmatrix}\begin{pmatrix}-2\\4\end{pmatrix}+\frac{1}{14}\begin{pmatrix}9c_1-6c_2-3c_3\\-6c_1+4c_2+2c_3\\-3c_1+2c_2+\ \ c_3\end{pmatrix}$$

其中 $Y=\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix}$为$\mathbb{F}^3$ 中的任意元素.

□

以下分别介绍 A 的$\{1,4\}$-逆与极小范数解,$\{1,3\}$-逆与最小二乘解,A^{+}与极小最小二乘解之间的关系.

(Ⅰ)*A 的$\{1,4\}$-逆与极小范数解*

对于相容方程组 $AX=b$,由定理 3.3.1 可知,只需随意找到 A 的一个$\{1\}$-逆 A^{-},该方程组的通解就可表示成:

$$X=A^{-}b+(I-A^{-}A)Y,(Y\in\mathbb{F}^n)$$

现在这些解中找一个解 X,使其范数最小,即在相容方程组 $AX=b$ 的解集中找出一个X_0,使得对 $AX=b$ 的任何解X,均有 $\|X_0\|\leqslant\|X\|$,这里 $\|\cdot\|$ 表示欧几里得范数.这样的 X_0 称为相容方程组 $AX=b$ 的一个**极小范数解**.显然,对一个相容的 $AX=b$,首要的问题是:其极小范数解存在否?进一步,倘若存在,它是否唯一呢?答案是肯定的.下面的推导表明:A 的$\{1,4\}$-逆在解决此问题中发挥着关键的作用.

定理 3.3.4 相容线性方程 $AX=b$ 的极小范数解存在且唯一,它为 Gb, $G\in A\{1,4\}$.

为证此定理,需要如下引理.

引理 3.3.1 设 $G\in A\{1\}$,以下三条等价:

(1)对任意使得 $AX=b$ 相容的 b, Gb 均是 $AX=b$ 的一个极小范数解;

(2)对任意的 $Y,Z\in\mathbb{F}^n$, G 满足 $Z^*(GA)^*(I-GA)^*Y=0$;

(3)$(GA)^*(I-GA)=0$.

证明: (2)⇔(3) 是显然的.

(1)⇒(2) $\forall Z\in\mathbb{F}^n$,令 $b=AZ$,于是 $AX=b$ 构成一个相容线性方程组,并且其通解为:

$$X=GAZ+(I-GA)Y,(Y\in\mathbb{F}^n)$$

结果:

$$\begin{aligned}\|X\|^2&=\langle GAZ+(I-GA)Y,GAZ+(I-GA)Y\rangle\\&=\|GAZ\|^2+\|(I-GA)Y\|^2+2\mathrm{Re}\langle(I-GA)Y,GAZ\rangle\\&=\|GAZ\|^2+\|(I-GA)Y\|^2+2\mathrm{Re}(Z^*A^*G^*(I-GA)Y),\end{aligned}$$

其中 Y 为自由变量.

现在由(1)知, $\|GAZ\|^2=\|Gb\|^2\leqslant\|X\|^2$,结果

$$\|(I-GA)Y\|^2+2\mathrm{Re}(Z^*A^*G^*(I-GA)Y)\geqslant 0 \tag{3.3.1}$$

其中 Y,Z 均为$\mathbb{F}^n$ 中任意元.断言:

对任何 $Y,Z\in\mathbb{F}^n$,必有 $Z^*A^*G^*(I-GA)Y=0$.否则,$\exists Y_0,Z_0\in\mathbb{F}^n$,有 $Z_0^*A^*G^*(I-GA)Y_0\neq0$.于是要么 $Z_0^*(GA)^*(I-GA)Y_0$ 的实部 Re 不为 0,要么 $Z_0^*(GA)^*(I-GA)Y_0$ 的虚部 Im 不为 0. 当 $Z_0^*(GA)^*(I-GA)Y_0$ 的虚部 Im 不为 0 时,取 $Y_1=\mathrm{i}Y_0$,则二元对(Y_1,Z_0)满足: $Z_0^*(GA)^*(I-GA)Y_1$ 的实部不为 0.故现可假定 $Z_0^*(GA)^*(I-GA)Y_0$ 的实部 Re 不为 0.由于 $\|(I-GA)Y_0\|^2$ 是一个固定的数,故当 $\mathrm{Re}(Z_0^*(GA)^*(I-GA)Y_0)$为负数时,取 k 为足够大的正数就有:

$$\|(I-GA)Y_0\|^2+2\mathrm{Re}((kZ_0)^*(GA)^*(I-GA)Y_0)<0$$

与(3.3.1)式矛盾,当 $\mathrm{Re}(Z_0^*(GA)^*(I-GA)Y_0)$为正数时,取 k 为模足够大的负数,仍有

$$\|(I-GA)Y_0\|^2+2\mathrm{Re}((kZ_0)^*(GA)^*(I-GA)Y_0)<0$$

仍与(3.3.1)式矛盾,于是断言真,从而(1)⇒(2)获证.

(2)⇒(1) 由于 $AX=b$ 相容, b 可以表示为 $b=AZ$ 的形式,同时 $AX=b$ 的任何解 X 可表示为

$$X=GAZ+(I-GA)Y$$

因此,由等式 $\|X\|^2=\|GAZ\|^2+\|(I-GA)Y\|^2+2\mathrm{Re}(Z^*A^*G^*(I-GA)Y)$立即知: $\|X\|^2\geqslant\|Gb\|^2$.

□

引理 3.3.2 设 $G\in\mathbb{F}^{n\times m},A\in\mathbb{F}^{m\times n}$,以下三条等价:

(1)对任意使得 $AX=b$ 相容的 b, Gb 均是 $AX=b$ 的一个极小范数解;

(2)$G\in A\{1,4\}$,即 G 满足$\begin{cases}AGA=A\\(GA)^*=GA\end{cases}$;

(3)$GAA^*=A^*$.

证明：(1)⇒(2) 由(1)知，$\forall Z\in \mathbb{F}^{n}$，有 GAZ 是 $AX=AZ$ 的极小范数解，故有 $AGAZ=AZ$．由 Z 的任意性知，$AGA=A$；另一方面，由引理3.3.1的(3)知：$(GA)^{*}(I-GA)=0$，从而 $(GA)^{*}=(GA)^{*}GA$ 为自伴阵，结果 $(GA)^{*}=GA$．

(2)⇒(1) 直接验证得 $(GA)^{*}(I-GA)=0$，故由引理3.3.1即得．

(2)⇒(3) $GAA^{*}=(GA)^{*}A^{*}=(AGA)^{*}=A^{*}$．

(3)⇒(2) $GAA^{*}=A^{*}$ 导致 $AGAA^{*}=AA^{*}$，从而由第1章的习题3知 $AGA=A$．此外，不难验证 $(GA)^{*}=GA$．

□

定理3.3.4的证明 (1)存在性：由 A 的$\{1,4\}$－逆总存在，故由引理3.3.2知相容方程组 $AX=b$ 的极小范数解必存在．

(2)唯一性：取定一个 $G\in A\{1,4\}$，X_1 也是相容线性方程组 $AX=b$ 的极小范数解，则由定理3.3.1知 X_1 可表为

$$X_1=Gb+(I-GA)Y_1$$

结果由引理3.3.1及引理3.3.2得：

$$\|X_1\|^2=\|Gb\|^2+\|(I-GA)Y_1\|^2=\|Gb\|^2$$

故

$$(I-GA)Y_1=0,X_1=Gb$$

□

至此可看出，给定一个相容线性方程组 $AX=b$，其通解可用 A 的任何一个$\{1\}$－逆借助于定理3.3.1将其给出；其极小范数解，则不但存在且唯一，并且可用 A 的任何一个$\{1,4\}$－逆 G 将其表示出，它就是 Gb．这也正是我们把 A 的$\{1,4\}$－逆叫做 A 的极小范数 g 逆的原因．

(Ⅱ)A 的$\{1,3\}$－逆与最小二乘解

给定线性方程组 $AX=b$，$(A\in \mathbb{F}^{m\times n},b\in \mathbb{F}^{m},X\in \mathbb{F}^{n})$，当它不相容时，通常退而求其次，转而求其在某种意义下误差最小的近似解．

定义3.3.1 线性方程组 $AX=b$ 的一个最小二乘解 X_0 是指：X_0 为满足条件

$$\|AX_0-b\|=\min\{\|AX-b\|\mid X\in \mathbb{F}^{n}\}$$

的向量．

如同前面所述的相容线性方程组的极小范数解情形，首先要解决的是最小二乘解是否对任何一个线性方程组皆存在的问题．进一步，在存在的情形下，其是否唯一？继而它的一般表达式又如何？通过以下的推导发现，利用矩阵的$\{1,3\}$－逆能很好地解决上述系列问题．

定理3.3.5 设 $G\in \mathbb{F}^{n\times m}$，$A\in \mathbb{F}^{m\times n}$，以下三条等价：

(1)对 $\forall b\in \mathbb{F}^{m}$，$Gb$ 均是 $AX=b$ 的最小二乘解；

(2) $G\in A\{1,3\}$ 即，G 满足 $\begin{cases}AGA=A\\(AG)^{*}=AG\end{cases}$；

(3) $A^{*}AG=A^{*}$．

证明：(1)⇒(2)　$\forall X\in \mathbb{F}^n$，考虑 $\|AX-b\|^2$.

$$\|AX-b\|^2 = \|AGb-b+A(X-Gb)\|^2 = \|AGb-b\|^2+\|A(X-Gb)\|^2+2\mathrm{Re}((AGb-b)^*A(X-Gb))$$

于是

$$\|AX-b\|^2-\|AGb-b\|^2=\|A(X-Gb)\|^2+2\mathrm{Re}((AGb-b)^*A(X-Gb))$$

结果有：

$$\forall b\in \mathbb{F}^m,\forall X\in \mathbb{F}^n,\|A(X-Gb)\|^2+2\mathrm{Re}((AGb-b)^*A(X-Gb))\geqslant 0$$

令 $Y=X-Gb$，于是有

$$\|AY\|^2+2\mathrm{Re}(b^*(AG-I)^*AY)\geqslant 0$$

由 b,X 的任意性及(1)知

$$\forall b\in \mathbb{F}^m,\forall Y\in \mathbb{F}^n,\text{恒有 }\|AY\|^2+2\mathrm{Re}(b^*(AG-I)^*AY)\geqslant 0$$

此时断言：$\forall b\in \mathbb{F}^m,\forall Y\in \mathbb{F}^n$，必有 $b^*(AG-I)^*AY=0$，从而 $(AG-I)^*A=0$，即

$$(AG)^*A=A$$

事实上，若 $\exists b_0\in \mathbb{F}^m,Y_0\in \mathbb{F}^n$ 使得 $b_0^*(AG-I)^*AY_0\neq 0$，于是可设

$$\mathrm{Re}(b_0^*(AG-I)^*AY_0)\neq 0$$

取足够大的正数 k(或者足够小的负数 k)后有

$$\|AY_0\|^2+2\mathrm{Re}(b_0^*(AG-I)^*AkY_0)<0$$

矛盾，断言真.

现由式子 $(AG)^*A=A$，两边右乘以 G，得 $(AG)^*AG=AG$，于是 AG 为自伴阵.再次观察等式 $(AG)^*A=A$，它变为 $AGA=A$.至此，$G\in A\{1,3\}$.

(2)⇒(1)　已知 $G\in A\{1,3\}$，于是 $\forall b\in \mathbb{F}^m,\forall X\in \mathbb{F}^n$，有

$$\|AX-b\|^2-\|AGb-b\|^2=\|A(X-Gb)\|^2+2\mathrm{Re}((AGb-b)^*A(X-Gb)) = \|A(X-Gb)\|^2+0\geqslant 0$$

因此 Gb 为 $AX=b$ 的最小二乘解.

(2)⇔(3)　显然.结合矩阵$\{1,3\}$－逆的存在性，现已经证明了.

□

定理 3.3.6　任何线性方程组 $AX=b$，其最小二乘解必存在，且任取 $A\{1,3\}$ 中的元 G，Gb 即为 $AX=b$ 的一个最小二乘解.

正是基于此定理，人们常常把 A 的$\{1,3\}$－逆称作为 A 的最小二乘 g 逆.

值得指出的是，线性方程组的最小二乘解可能不唯一.事实上，对相容的 $AX=b$ 而言，只要其解不唯一，则该方程组的最小二乘解就不唯一.对不相容线性方程组 $AX=b$ 而言，其最小二乘解同样可以不唯一，除非 A 的列向量线性无关.

例 3.3.3　求不相容方程组

$$\begin{cases}x_1+2x_2+3x_3=1\\x_1+x_3=0\\2x_1+2x_3=1\\2x_1+4x_2+6x_3=3\end{cases}$$

的最小二乘解.

解　方法一　将方程组写成矩阵形式 $AX=b$, $A=\begin{pmatrix}1&2&3\\1&0&1\\2&0&2\\2&4&6\end{pmatrix}$, $X=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$, $b=\begin{pmatrix}1\\0\\1\\3\end{pmatrix}$.

于是 $\mathrm{rank}A=2$, $\mathrm{rank}(A,b)=3$,该方程组为不相容方程组.

对 A 进行初等变换得:

$$PAQ=A_1=\begin{pmatrix}1&2&0\\0&-2&0\\0&0&0\\0&0&0\end{pmatrix}=\begin{pmatrix}A_2&0\\0&0\end{pmatrix}$$

其中 $P=\begin{pmatrix}1&0&0&0\\-1&1&0&0\\0&-2&1&0\\-2&0&0&1\end{pmatrix}$, $Q=\begin{pmatrix}1&0&-1\\0&1&-1\\0&0&1\end{pmatrix}$, $A_2=\begin{pmatrix}1&2\\0&-2\end{pmatrix}$.再求得

$$P^{-1}=\begin{pmatrix}1&0&0&0\\1&1&0&0\\2&2&1&0\\2&0&0&1\end{pmatrix},\qquad Q^{-1}=\begin{pmatrix}1&0&1\\0&1&1\\0&0&1\end{pmatrix}$$

$$C=P^{-1}\begin{pmatrix}A_2\\0\end{pmatrix}=\begin{pmatrix}1&2\\1&0\\2&0\\2&4\end{pmatrix},\qquad D=(I\quad 0)Q^{-1}=\begin{pmatrix}1&0&1\\0&1&1\end{pmatrix}$$

$$C_L^{-1}=(C^*C)^{-1}C^*=\frac{1}{10}\begin{pmatrix}0&2&4&0\\1&-1&-2&2\end{pmatrix}$$

$$D_R^{-1}=D^*(DD^*)^{-1}=\frac{1}{3}\begin{pmatrix}2&-1\\-1&2\\1&1\end{pmatrix}$$

结果 $A_l^-=D_R^{-1}C_L^{-1}=\frac{1}{30}\begin{pmatrix}-1&5&10&-2\\2&-4&-8&4\\1&1&2&2\end{pmatrix}$.因此最小二乘解为

$$X=A_l^-b=\frac{1}{30}\begin{pmatrix}-1&5&10&-2\\2&-4&-8&4\\1&1&2&2\end{pmatrix}\begin{pmatrix}1\\0\\1\\3\end{pmatrix}=\frac{1}{10}\begin{pmatrix}1\\2\\3\end{pmatrix}$$

经计算,其最小误差平方和为

$$\|AA_l^-b-b\|^2=\frac{10}{25}=\frac{2}{5}$$

方法二　通过计算，$A^*A=10\begin{pmatrix}1&1&2\\1&2&3\\2&3&5\end{pmatrix}$，对 A^*A 进行初等行变换，化为

$$P(A^*A)=\begin{pmatrix}10&0&10\\0&10&10\\0&0&0\end{pmatrix}=\begin{pmatrix}C\\0\end{pmatrix}$$

其中 $P=\begin{pmatrix}2&-1&0\\-1&1&0\\-1&-1&1\end{pmatrix}$，$C=\begin{pmatrix}10&0&10\\0&10&10\end{pmatrix}$，于是

$$C_R^{-1}=C^*(CC^*)^{-1}=\begin{pmatrix}10&0\\0&10\\10&10\end{pmatrix}\begin{pmatrix}2&1\\1&2\end{pmatrix}^{-1}\frac{1}{100}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}$$

$$=\begin{pmatrix}10&0\\0&10\\10&10\end{pmatrix}\frac{1}{3}\cdot\frac{1}{100}\begin{pmatrix}2&-1\\-1&2\end{pmatrix}=\frac{1}{30}\begin{pmatrix}2&-1\\-1&2\\1&1\end{pmatrix}$$

结果 $(C_R^{-1},0)P=\frac{1}{30}\begin{pmatrix}2&-1\\-1&2\\1&1\end{pmatrix}\begin{pmatrix}2&-1&0\\-1&1&0\\-1&-1&1\end{pmatrix}=\frac{1}{30}\begin{pmatrix}5&-3&0\\-4&3&0\\1&0&0\end{pmatrix}$ 是 A^*A 的一个 $\{1\}$ – 逆.因此

$$\frac{1}{30}\begin{pmatrix}5&-3&0\\-4&3&0\\1&0&0\end{pmatrix}\cdot A^*b=\frac{1}{30}\begin{pmatrix}5&-3&0\\-4&3&0\\1&0&0\end{pmatrix}\begin{pmatrix}1&1&2&2\\2&0&0&4\\3&1&2&6\end{pmatrix}\begin{pmatrix}1\\0\\1\\3\end{pmatrix}$$

$$=\frac{1}{30}\begin{pmatrix}5&-3&0\\-4&3&0\\1&0&0\end{pmatrix}\begin{pmatrix}9\\14\\23\end{pmatrix}=\frac{1}{10}\begin{pmatrix}1\\2\\3\end{pmatrix}$$

是所求方程的最小二乘解.

□

推论 3.3.1　线性方程组 $AX=b$ 的最小二乘解的通式为

$$X=A_l^-b+(I_n-A_l^-A)Y,\ Y\in\mathbb{F}^n$$

其中 A_l^- 为 A 的任意取定的一个 $\{1,3\}$ – 逆.进一步，最小二乘解唯一当且仅当 A 为列满秩阵.

证明： 首先，直接可知，$X=A_l^-b+(I_n-A_l^-A)Y$ 是 $AX=b$ 的最小二乘解.

其次，若 X_0 为 $AX=b$ 最小二乘解，则有 X_0、A_l^-b 均为 $AX=b$ 的最小二乘解，故有 $\|AX_0-b\|=\|AA_l^-b-b\|$.结果

$$\begin{aligned}0&=\|AX_0-b\|^2-\|AA_l^-b-b\|^2\\&=\|A(X_0-A_l^-b)\|^2+2\mathrm{Re}[b^*(AA_l^--I)^*A(X_0-A_l^-b)]\end{aligned}$$

$$= \| A(X_0 - A_l^- b) \|^2 + 0$$

因此 $A(X_0 - A_l^- b) = 0$，$X_0 - A_l^- b$ 为 $AX = 0$ 的一个解，由 $AX = 0$ 的通解公式知：

$$\exists\, Y_0 \in \mathbb{F}^n, \text{s.t.}\, X_0 - A_l^- b = (I - A_l^- A) Y_0$$

移项即有

$$X_0 = A_l^- b + (I - A_l^- A) Y_0$$

最后，由定理 3.2.2 知，A 列满秩 $\Leftrightarrow$ A 的 $\{1,3\}$ – 逆唯一（它就是 $(A^* A)^{-1} A^*$）. 因此当 A 列满秩时，$AX = b$ 的最小二乘解的通式为 $X = A_l^- b$，唯一.

反过来，设 $AX = b$ 的最小二乘解唯一. 若 A 不为列满秩，则 A 没有左逆，从而任取 A_l^-，$I_n - A_l^- A \neq 0$，更可选取 $Y \in \mathbb{F}^n$ 满足 $(I_n - A_l^- A) Y \neq 0$. 此时

$$A_l^- b \neq A_l^- b + (I_n - A_l^- A) Y$$

而 $A_l^- b$，$A_l^- b + (I_n - A_l^- A) Y$ 均是 $AX = b$ 的最小二乘解，与假设矛盾，推论获证.

□

下面考察矩阵的加号逆（Moore – Penrose 逆）在线性方程组求解中的应用.

（Ⅲ）A 的加号逆与极小最小二乘解

给定一个线性方程组 $AX = b$ $(A \in \mathbb{F}^{m\times n}, X \in \mathbb{F}^n, b \in \mathbb{F}^m)$，由推论 3.3.1 知，该方程组的最小二乘解的集合为 $T = \{A_l^- b + (I_n - A_l^- A) Y \mid Y \in \mathbb{F}^n\}$. 现在想从 T 中找出一个范数最小的，称之为线性方程组 $AX = b$ 的**极小最小二乘解**.

定理 3.3.7 线性方程组 $AX = b$ 的最小二乘解与相容方程组 $A^* AX = A^* b$ 的解一致，且还与相容方程组 $AX = AA_l^- b$ 的解一致，即

$$\{X \mid A^* AX = A^* b\} = \{X \mid \|AX - b\|_2 = \min\} = \{X \mid AX = AA_l^- b\}$$

证明：(1) 显然 $X = A_l^- b$ 为方程 $AX = AA_l^- b$ 的解，从而 $AX = AA_l^- b$ 为相容方程组.

$\forall X \in \{X \mid AX = AA_l^- b\}$，$A^* AX = A^* (AA_l^- b) = (A^* AA_l^-) b = A^* b$（定理 3.3.5）.

因此 X 是 $A^* AX = A^* b$ 的解，$\{X \mid AX = AA_l^- b\} \subseteq \{X \mid A^* AX = A^* b\}$.

(2) 由 $A^* AA_l^- b = A^* b$ 知 $A^* AX = A^* b$ 相容. 任取 Y 是它的解，则

$$\begin{aligned}
\| AX - b \|^2 &= \| A(X - Y) + (AY - b) \|^2 \\
&= \| A(X - Y) \|^2 + \| AY - b \|^2 + 2\mathrm{Re}(X - Y)^* A^* (AY - b) \\
&= \| A(X - Y) \|^2 + \| AY - b \|^2 \geqslant \| AY - b \|^2
\end{aligned}$$

上式说明 Y 是方程 $AX = b$ 的最小二乘解. 即

$$\{X \mid A^* AX = A^* b\} \subseteq \{X \mid \|AX - b\|_2 = \min\}$$

(3) 由 (2) 知方程组 $AX = b$ 的最小二乘解存在. 任取 X 为它的最小二乘解，由 (1)、(2) 知 $A_l^- b \in \{X \mid AX = AA_l^- b\} \subseteq \{X \mid A^* AX = A^* b\} \subseteq \{X \mid \|AX - b\|_2 = \min\}$，由此 $A_l^- b$ 也是最小二乘解. 故有 $\| AX - b \| = \| AA_l^- b - b \| = \min$. 由于

$$\begin{aligned}
0 &= \| AX - b \|^2 - \| AA_l^- b - b \|^2 \\
&= \| A(X - A_l^- b) + (AA_l^- b - b) \|^2 - \| AA_l^- b - b \|^2 \\
&= \| A(X - A_l^- b) \|^2 + 2\mathrm{Re}[A(X - A_l^- b)]^* (AA_l^- b - b) \\
&= \| A(X - A_l^- b) \|^2 + 2\mathrm{Re}(X - A_l^- b)^* A^* (AA_l^- - I) b
\end{aligned}$$

$$= \| A(X - A_l^- b) \|^2$$

故 $A(X - A_l^- b) = 0, AX = AA_l^- b$.结果

$$\{X \mid A^* AX = A^* b\} = \{X \mid \|AX - b\| = \min\} = \{X \mid AX = AA_l^- b\}$$

□

定理 3.3.8　线性方程组 $AX = b$ 的极小最小二乘解存在且唯一,它为 $X = A^+ b$.

证明: 由定理 3.3.7 知,$\{Y \mid \|AY - b\| = \min\} = \{Y \mid AY = AA_l^- b\}$,于是

$$\begin{aligned}
&\{X \mid \|X\| = \min, X \in \{Y \mid \|AY - b\| = \min\}\} \\
=& \{X \mid \|X\| = \min, X \in \{Y \mid AY = AA_l^- b\}\} \\
=& \{X \mid \|X\| = \min, AX = AA_l^- b\} \\
=& \{A_m^- (AA_l^- b)\} \\
=& \{A^+ b\}
\end{aligned}$$

□

矩阵的加号逆还有如下基本性质.

定理 3.3.9　设 $A \in \mathbb{F}^{m \times n}$.则有

(1) $\operatorname{rank} A = \operatorname{rank} A^+ = \operatorname{rank} A^+ A = \operatorname{rank} AA^+ = m - \operatorname{rank}(I_m - AA^+) = n - \operatorname{rank}(I_n - A^+ A)$;

(2) $(A^+)^+ = A$;

(3) $(A^+)^* = (A^*)^+$;

(4) $(\lambda A)^+ = \frac{1}{\lambda} A^+, \lambda \in \mathbb{F}, \lambda \neq 0$;

(5) $A = AA^* (A^*)^+ = (A^*)^+ A^* A$;

(5′) $A^* = A^* AA^+ = A^+ AA^*$;

(6) $(A^* A)^+ = A^+ (A^*)^+$;

(7) $A^+ = (A^* A)^+ A^* = A^* (AA^*)^+$;

(8) $A^+ AB = A^+ AC \Leftrightarrow AB = AC$;

(9) $(UAV)^+ = V^* A^+ U^*$,其中 $U^* U = I_m, V^* V = I_n$.

证明: 留给读者作为练习.

□

例 3.3.4　求方程组 $\begin{cases} x_1 + 2x_2 = 1 \\ x_3 = 1 \\ 2x_1 + 4x_2 = 3 \end{cases}$ 的极小最小二乘解.

解: 由方程组知

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 2 & 4 & 0 \end{pmatrix}, \ b = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

于是 $\operatorname{rank} A = 2, \operatorname{rank}(A, b) = 3$.

现求出 A^+.通过计算知 $A^+ = \frac{1}{25}\begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \\ 0 & 25 & 0 \end{pmatrix}$.结果,所求的极小最小二乘解为:

$$X = A^+ b = \frac{1}{25}\begin{pmatrix} 7 \\ 14 \\ 25 \end{pmatrix}$$

□

习题三

1. 设 $A \in \mathbb{F}^{m\times n}$,证明 A 列满秩$\Leftrightarrow A$ 的$\{1,3\}$ - 逆唯一．进一步,此时它就是$(A^* A)^{-1}A^*$.

2. 举例说明 $\text{rank}A^- > \text{rank}A$ 有可能成立.

3. 证明下面三个结论中的任何两个都能推导出第三个:

(1)$G \in A\{1\}$;

(2)$G \in A\{2\}$;

(3)$\text{rank}\,G = \text{rank}A$.

4. 设 $A \in \mathbb{C}_r^{m\times n}$,$r > 0$,证明对任意指定的正整数 $s \in [0, r]$,必存在 $X_s \in A\{2\}$使$\text{rank}\,X_s = s$.

5. 设 $A \in \mathbb{C}_r^{m\times n}$($r > 0$)的奇异值分解为

$$A = U\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^*$$

其中$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}_{r\times r}$,$\sigma_1 \geqslant \cdots \geqslant \sigma_r$.证明:$A^+ = V\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^*$.

6. 设 $A \in \mathbb{C}^{m\times n}$,证明:$A^+ = (A^* A)^+ A^* = A^*(AA^*)^+$.

7. 证明:矩阵方程组$\begin{cases} AX = B \\ XD = E \end{cases}$有解的充要条件是 $AX = B$ 与 $XD = E$ 分别有解,且 $AE = BD$.

8. 设 $A \in \mathbb{C}^{m\times n}$,证明:下面最小问题$\begin{cases} \| AXB - C \|_F = \min \\ \| X \|_F = \min \end{cases}$的解为 $X = A^+ CB^+$.

9. 证明定理 3.3.9 中除(7)以外的其他各条.

10. 设 $A \in \mathbb{C}^{m\times n}$,$X \in \mathbb{C}^{n\times m}$. 证明:$X \in A\{1,3\} \Leftrightarrow A^* AX = A^*$.

第 4 章　矩阵范数与矩阵级数

在数学的许多分支和工程实际中，特别涉及多元分析时，仅仅研究矩阵代数是不够的，还必须研究常数矩阵序列和级数的收敛性．本章内容的安排是：先介绍矩阵范数，然后讨论矩阵序列与矩阵级数的收敛性，从而给出矩阵级数的 Lagrange-Sylvester 定理，最后作为矩阵序列与矩阵级数的应用，给出非负矩阵论中著名的 Perron 定理的完整证明以及给出计算 A^+ 的几种方法．

4.1　矩阵范数

数域$\mathbb{F}$上的向量空间${}_{\mathbb{F}}V$，当引进满足下列三条件的函数 $\|\cdot\|: {}_{\mathbb{F}}V\to\mathbb{R}$ 后便构成所谓的**赋范空间**，其中 $\|\cdot\|$ 称为**范数**．这三个条件分别是：

(1)（正定性）$\|x\|\geqslant 0$ 且 $\|x\|=0\Leftrightarrow x=0,\ \forall x\in V$；

(2)（齐次性）$\|kx\|=|k|\,\|x\|,\ \forall k\in F,\ \forall x\in V$；

(3)（三角不等式）$\|x+y\|\leqslant\|x\|+\|y\|,\ \forall x,y\in V$．

例如，在 n 维向量空间 $\mathbb{F}^n$ 中，设 $x=(x_1,\cdots,x_n)^{\mathrm{T}}$，使

$$\|x\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}\ (p\geqslant 1)$$

则 $\|\cdot\|_p$ 构成了$\mathbb{F}^n$ 的范数，称之为 **Hölder 范数**或 **p – 范数**．特别地，$\|\cdot\|_1$ 叫做$\mathbb{F}^n$ 的 **1 – 范数**，$\|\cdot\|_2$ 叫做$\mathbb{F}^n$ 的 2 – 范数或 **Euclid 范数**．同样 $\|x\|_\infty=\max\limits_{1\leqslant i\leqslant n}|x_i|$ 也给出了$\mathbb{F}^n$ 的一个范数，称该范数为 ∞ – 范数．

容易知道，$\mathbb{F}^{m\times n}$可看成为以 $\{E_{ij}\mid i=1,2,\cdots,m,j=1,2,\cdots,n\}$ 为基的 nm 维向量空间．现对该空间赋予某种范数 $\|\cdot\|$，则由于该空间的向量实际即为 $m\times n$ 阵，因此此时范数 $\|\cdot\|$ 又叫做**矩阵范数**．换句话说，当赋范空间的向量是矩阵时，称该空间的向量范数为矩阵范数．

例 4.1.1　对$\mathbb{F}^{m\times n}$中任何矩阵 $A=(a_{ij})_{m\times n}$，分别使

$$\|A\|_F=\left(\sum_{i,j}|a_{ij}|^2\right)^{\frac{1}{2}}$$

$$\|A\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|,\quad \|A\|_{v_1}=\sum_{i,j}|a_{ij}|$$

和

$$\| A \|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} | a_{ij} |$$

则 $\|\cdot\|_F$, $\|\cdot\|_1$, $\|\cdot\|_{v_1}$ 和 $\|\cdot\|_\infty$ 均为 $\mathbb{F}^{m\times n}$ 的矩阵范数.其中 $\|\cdot\|_F$ 也叫矩阵 A 的 **Frobenius** 范数或 Euclid 范数; $\|\cdot\|_1$ 又叫矩阵的 1 – 范数或**列范数**,而 $\|\cdot\|_\infty$ 又叫矩阵的 ∞ – 范数或**行范数**.

例 4.1.2 设 $\|\cdot\|$ 为 $\mathbb{F}^{l\times n}$ 上的矩阵范数,固定一个列满秩阵 $A\in\mathbb{F}^{l\times m}$, $\forall X\in \mathbb{F}^{m\times n}$,使

$$\| X \|' = \| AX \|$$

则 $\| X \|'$ 为 $\mathbb{F}^{m\times n}$ 上的矩阵范数.

上例表明,可以用已知的矩阵范数去构造新的矩阵范数.

例 4.1.3 $\forall A,B\in\mathbb{F}^{m\times n}$,使 $\langle A,B\rangle = \mathrm{tr}(AB^*)$,则有 $(\mathbb{F}^{m\times n},\langle\cdot\rangle)$ 构成一个内积空间.令 $\| A \| = \sqrt{\mathrm{tr}(AA^*)}$,则有 $\|\cdot\|$ 为 $\mathbb{F}^{m\times n}$ 的矩阵范数,它就是 $\| A \|_F$.一般地,给定一个实或复赋范空间 $(V,\|\cdot\|)$,范数 $\|\cdot\|$ 能被某个内积导出,即存在 V 上的某内积使 $\| x \| = \sqrt{\langle x,x\rangle}$ 对任意 $x\in V$ 皆成立,其充要条件是:范数满足所谓的**平行四边形法则**

$$\| x+y \|^2 + \| x-y \|^2 = 2\| x \|^2 + 2\| y \|^2$$

给定一个线性空间,人们可用多种不同的方式定义出多种不同的范数,为刻画这些范数之间的联系,一个重要的手段是范数等价.

定义 4.1.1 设 V 是线性空间, $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 均为 V 上的范数.若存在正数 M,m 使得对一切 $x\in V$ 满足

$$m\| x \|_\beta \leq \| x \|_\alpha \leq M\| x \|_\beta$$

则称 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ **是等价的**.

显然,范数的等价满足自反性,对称性和传递性.

定理 4.1.1 设 V 是实数域或复数域的有限维向量空间,则 V 的任何两种范数均是等价的.

证明:设 $\|\cdot\|_\alpha$ 为 V 的一种范数.取定 V 的一组基 $\{e_1,\cdots,e_n\}$.于是 $\forall x\in V$, $x = x_1e_1+\cdots+x_ne_n$.定义函数 $\varphi_\alpha:\mathbb{F}^n\to\mathbb{R}$, $(x_1,\cdots,x_n)\to\| x_1e_1+\cdots+x_ne_n \|_\alpha$ 则有 $\varphi_\alpha(x_1,\cdots,x_n) = \| x_1e_1+\cdots+x_ne_n \|_\alpha$ 为连续函数.事实上

$$\begin{aligned} |\varphi_\alpha(x_1,\cdots,x_n)-\varphi_\alpha(y_1,\cdots,y_n)| &\leq \| (x_1-y_1)e_1+\cdots+(x_n-y_n)e_n \|_\alpha \\ &\leq |x_1-y_1|\,\| e_1 \|_\alpha+\cdots+|x_n-y_n|\,\| e_n \|_\alpha \\ &\underset{\text{Hölder不等式}}{\leq} \Big(\sum_{i=1}^{n}|x_i-y_i|^2\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^{n}\| e_i \|_\alpha^2\Big)^{\frac{1}{2}} \end{aligned}$$

因此, φ_α 是关于 $x_1,\cdots,x_n$ 的多元连续函数.现假定 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 为 V 上的两种范数.考虑函数

$$f(x_1,\cdots,x_n):\mathbb{F}^n\to\mathbb{R}$$

$$(x_1,\cdots,x_n)\to\frac{\| x_1e_1+\cdots+x_ne_n \|_\alpha}{\| x_1e_1+\cdots+x_ne_n \|_\beta}$$

于是$f(x_1,\cdots,x_n)=\dfrac{\varphi_\alpha(x_1,\cdots,x_n)}{\varphi_\beta(x_1,\cdots,x_n)}$在其定义域内是连续函数.取单位球面

$$S=\{(\xi_1,\cdots,\xi_n):|\xi_1|^2+\cdots+|\xi_n|^2=1\}$$

则有S为$\mathbb{F}^n$上的有界闭集.因此f在S上有最大值M,也有最小值m.显然M,m均大于零.于是

$$\forall x\neq 0\in V,x=x_1e_1+\cdots+x_ne_n,\left(\frac{x_1}{\sqrt{|x_1|^2+\cdots+|x_n|^2}},\cdots,\frac{x_n}{\sqrt{|x_1|^2+\cdots+|x_n|^2}}\right)\in S$$

再令

$$\xi=\frac{x_1}{\sqrt{|x_1|^2+\cdots+|x_n|^2}}e_1+\cdots+\frac{x_n}{\sqrt{|x_1|^2+\cdots+|x_n|^2}}e_n$$

结果

$$m\leqslant\frac{\|\xi\|_\alpha}{\|\xi\|_\beta}=\frac{\|x\|_\alpha}{\|x\|_\beta}\leqslant M$$

亦即

$$m\|x\|_\beta\leqslant\|x\|_\alpha\leqslant M\|x\|_\beta$$

容易看出,上述不等式在$x=0$时是显然成立的.于是定理获证.

□

4.2　矩阵的算子范数

对于任给的$A\in\mathbb{F}^{m\times n}$,它可看成从$\mathbb{F}^n$到$\mathbb{F}^m$的线性算子$\sigma,\sigma(e_1,\cdots,e_n)=(e_1,\cdots,e_m)A$.因此,可采用线性算子的算子范数定义方式引进矩阵的范数.

定义4.2.1　设$A\in\mathbb{C}^{m\times n},x\in\mathbb{C}^n$,$\|x\|_\alpha$与$\|Ax\|_\beta$分别是$\mathbb{C}^n$与$\mathbb{C}^m$中的范数,令

$$\|A\|=\sup_{x\neq 0}\frac{\|Ax\|_\beta}{\|x\|_\alpha}=\sup_{\|x\|_\alpha=1}\|Ax\|_\beta$$

则称$\|A\|$为矩阵A经向量范数$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$导出的算子范数,简称为A的**算子范数**.

对方阵$A\in\mathbb{C}^{n\times n}$而言,让$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$取相同的向量范数,此时矩阵$A$经向量范数$\|\cdot\|_\alpha$和$\|\cdot\|_\beta$所导出的算子范数常称作为**由$\|\cdot\|_\alpha$所导出的范数**.

由定义4.2.1容易看出,此时的$\|A\|$实际上也就是使$\|Ax\|_\beta\leqslant M\|x\|_\alpha$对一切$x$皆成立的最小非负数$M$.

现在一个自然的问题是,上一节中我们引进的矩阵的列范数(i.e. 1－范数),行范数(i.e. ∞－范数)是否均可看成算子范数呢?结论是肯定的.

定理4.2.1　设$A=(a_{ij})\in\mathbb{C}^{m\times n}$,则

(1) $\|A\|_1=\sup\limits_{\|x\|_1=1}\|Ax\|_1$,即矩阵的列范数是由向量的1－范数所导出的算子范数,这里向量的1－范数是指向量的分量绝对值之和.

(2) $\| A \|_{\infty} = \sup_{\| x \|_{\infty} = 1} \| Ax \|_{\infty}$,即矩阵的行范数是由向量的∞－范数所导出的算子范数,这里向量的∞－范数是指向量的分量绝对值的最大值.

(3) A 经向量的 Euclid 范数所导出的算子范数,常记作为 $\| A \|_2$, $\| A \|_2 = \sup_{\| x \|_2 = 1} \| Ax \|_2 = \sqrt{\lambda_1}$,其中 $\| x \|_2$ 表示向量的 Euclid 范数,λ_1 为 $A^* A$ 的最大特征值.

$\| A \|_2$ 也通常被称为矩阵 A 的**谱范数**或 **2－范数**.

证明: (1)设 $x = (x_1, \cdots, x_n)^T \in \mathbb{C}^{n \times 1}$,则有 $Ax = \left(\sum_{j=1}^{n} a_{1j} x_j, \cdots, \sum_{j=1}^{n} a_{mj} x_j \right)^T$,于是

$$\begin{aligned} \| Ax \|_1 &= \left| \sum_{j=1}^{n} a_{1j} x_j \right| + \cdots + \left| \sum_{j=1}^{n} a_{mj} x_j \right| = \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \\ &\leqslant \sum_{i=1}^{m} \sum_{j=1}^{n} | a_{ij} x_j | = \sum_{i,j} | a_{ij} | \, | x_j | \\ &\leqslant \left(\max_{1 \leqslant j \leqslant n} \left(\sum_{i=1}^{m} | a_{ij} | \right) \right) \left(\sum_{j=1}^{n} | x_j | \right) \end{aligned}$$

因此

$$\sup_{\| x \|_1 = 1} \| Ax \|_1 \leqslant \max_{1 \leqslant j \leqslant n} \left(\sum_{i=1}^{m} | a_{ij} | \right) = \| A \|_1$$

另一方面,设 A 的列范数为 $\sum_{i=1}^{m} \left| a_{ij_0} \right|$,则有取 $x_0 = e_{j_0}$,于是 $\| x_0 \|_1 = 1$,且 $Ax_0 = (a_{1j_0} \cdots a_{mj_0})^T$,故有

$$\sup_{\| x \|_1 = 1} \| Ax \|_1 \geqslant \| Ax_0 \|_1 = \sum_{i=1}^{m} \left| a_{ij_0} \right| = \| A \|_1$$

至此,有

$$\| A \|_1 = \sup_{\| x \|_1 = 1} \| Ax \|_1$$

获证.

(2)设 $x = (x_1, \cdots, x_n)^T \in \mathbb{C}^{n \times 1}$,则

$$\begin{aligned} \| Ax \|_{\infty} &= \max_{1 \leqslant i \leqslant m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \\ &\leqslant \max_{1 \leqslant i \leqslant m} \left(\sum_{j=1}^{n} | a_{ij} | \, | x_j | \right) \\ &\leqslant \left(\max_{1 \leqslant i \leqslant m} \left(\sum_{j=1}^{n} | a_{ij} | \right) \right) \max_{1 \leqslant j \leqslant n} | x_j | \end{aligned}$$

因此

$$\sup_{\| x \|_{\infty} = 1} \| Ax \|_{\infty} \leqslant \max_{1 \leqslant i \leqslant m} \left(\sum_{j=1}^{n} | a_{ij} | \right) = \| A \|_{\infty}$$

另一方面,设 $\| A \|_{\infty} = \sum_{j=1}^{n} \left| a_{i_0 j} \right| \neq 0$.取 $x_0 = (\xi_1, \cdots, \xi_n)^T$ 如下:

$$\xi_j = \begin{cases} \dfrac{|a_{i_0 j}|}{a_{i_0 j}}, 当\ a_{i_0 j} \neq 0 时 \\ 0, \quad 当\ a_{i_0 j} = 0 时 \end{cases}$$

于是

$$\| x_0 \|_\infty = 1$$

$$\left| \sum_{j=1}^{n} a_{ij}\xi_j \right| \leqslant \sum_{j=1}^{n} |a_{ij}| |\xi_j| \leqslant \sum_{j=1}^{n} |a_{i_0 j}|$$

这表明

$$\| Ax_0 \|_\infty = \max_{1\leqslant i\leqslant m} \left| \sum_{j=1}^{n} a_{ij}\xi_j \right| = \sum_{j=1}^{n} \left| a_{i_0 j} \right| = \| A \|_\infty$$

所以

$$\| A \|_\infty = \sup_{\| x \|_\infty = 1} \| Ax \|_\infty$$

(3)设 A^*A 的 n 个特征值为 $\lambda_1 \geqslant \lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$,其对应的正交单位特征向量为 $\xi_1, \xi_2, \cdots, \xi_n$.于是 $\forall$ 单位向量 $x \in \mathbb{C}^{n\times 1}$, x 可表示为 $x = a_1\xi_1 + \cdots + a_n\xi_n$,其中 $|a_1|^2 + \cdots + |a_n|^2 = 1$.结果

$$\begin{aligned}
\| Ax \|_2^2 &= \langle Ax, Ax \rangle = x^* A^* Ax \\
&= (\bar{a}_1 \xi_1^* + \cdots + \bar{a}_n \xi_n^*) A^* A (a_1\xi_1 + \cdots + a_n\xi_n) \\
&= (\bar{a}_1 \xi_1^* + \cdots + \bar{a}_n \xi_n^*)(a_1\lambda_1\xi_1 + \cdots + a_n\lambda_n\xi_n) \\
&= |a_1|^2 \lambda_1 + \cdots + |a_n|^2 \lambda_n \leqslant \lambda_1 \cdot 1 = \lambda_1
\end{aligned}$$

于是

$$\sup_{\| x \|_2 = 1} \| Ax \|_2 \leqslant \sqrt{\lambda_1}$$

另一方面,取 $x = \xi_1$,则 $\| \xi_1 \|_2 = 1$, $\| A\xi_1 \|_2^2 = \xi_1^* A^* A\xi_1 = \xi_1^* \lambda_1 \xi_1 = \lambda_1$.这表明

$$\sup_{\| x \|_2 = 1} \| Ax \|_2 \geqslant \sqrt{\lambda_1}$$

进而

$$\sup_{\| x \|_2 = 1} \| Ax \|_2 = \sqrt{\lambda_1}.$$

□

与矩阵的 1 - 范数,∞ - 范数相比较,矩阵的谱范数计算较为不便,但矩阵的谱范数具有一些特殊性质,使得它在理论和实际工程中有着广泛的用途.首先,有:

定理 4.2.2　设 $A \in \mathbb{F}^{m\times n}$, $\langle \cdot \rangle$ 为 $\mathbb{F}^m$ 上的通常内积,则有以下几条成立:

(1) $\| A \|_2 = \sup\limits_{x\neq 0} \dfrac{\| Ax \|_2}{\| x \|_2}$(谱范数的定义);

(1′) $\| A \|_2 = \sup\limits_{\| x \|_2 = 1} \| Ax \|_2$;

(2) $\| A \|_2 = \sup\limits_{x\neq 0, y\neq 0} \dfrac{|\langle Ax, y \rangle|}{\| x \|_2 \| y \|_2}$;

(3) $\| A \|_2 = \sup\limits_{\| x \|_2 = \| y \|_2 = 1} |\langle Ax, y \rangle|$;

(4) $\| UAV \|_2 = \| A \|_2$,对任何酉阵 U,V;

(5) $\| A^* \|_2 = \| A \|_2$.

证明: (1)和(1′)等价是显然的.(2)和(3)等价也是显然的.下证(1)⇔(2).

首先由许瓦兹不等式有

$$\frac{|\langle Ax, y\rangle|}{\| x \|_2 \| y \|_2} \leqslant \frac{\| Ax \|_2 \cdot \| y \|_2}{\| x \|_2 \cdot \| y \|_2} \leqslant \| A \|_2$$

因此

$$\sup_{x\neq 0, y\neq 0} \frac{|\langle Ax, y\rangle|}{\| x \|_2 \| y \|_2} \leqslant \| A \|_2$$

另一方面,取 $x_0 \in \mathbb{F}^n$ 使得 $\| x_0 \|_2 = 1$ 且 $\| A \|_2 = \| Ax_0 \|_2$,于是令 $y_0 = \frac{Ax_0}{\| Ax_0 \|_2}$,则当 $A \neq 0$ 时有

$$\| y_0 \|_2 = 1$$

$$|\langle Ax_0, y_0\rangle| = |y_0^* Ax_0| = \left| \frac{(Ax_0)^*}{\| Ax_0 \|_2} Ax_0 \right| = \| Ax_0 \|_2$$

因此又有

$$\sup_{\| x \|_2 = \| y \|_2 = 1} |\langle Ax, y\rangle| \geqslant \| Ax_0 \|_2 = \| A \|_2$$

结果无论 A 是否为零矩阵,均有

$$\sup_{x\neq 0, y\neq 0} \frac{|\langle Ax, y\rangle|}{\| x \|_2 \| y \|_2} = \| A \|_2$$

至此,(1),(1′),(2),(3)等价已获证.

关于(4),由谱范数的表达式知

$$\begin{aligned}\| UAV \|_2 &= \sqrt{\lambda_{\max}(V^* A^* U^* UAV)} \\ &= \sqrt{\lambda_{\max}(V^* A^* AV)} = \sqrt{\lambda_{\max}(A^* A)} = \| A \|_2\end{aligned}$$

对任意的酉阵 U,V 恒成立,因此(4)真.

注意到 $A^* A$ 的最大特征值与 AA^* 的最大特征值相同,因此得(5).

□

倘若给定的矩阵 $A \in \mathbb{F}^{n\times n}$ 为自伴(即自共轭)阵,则 A 的谱范数的计算还有下列更为简洁的形式.

定理 4.2.3 设 $A = (a_{ij}) \in \mathbb{F}^{n\times n}$ 为自共轭阵,$\langle\cdot\rangle$ 为 $\mathbb{F}^n$ 上的通常内积,令

$$\phi = \left\{ \frac{\langle Ax, x\rangle}{\langle x, x\rangle} : x \neq 0 \right\}$$

$$\varphi = \{\langle Ax, x\rangle : \| x \|_2 = 1\}$$

则有 $\phi = \varphi$, $\| A \|_2 = \max\{|\inf\phi|, |\sup\phi|\}$.

证明: (1) $\phi = \varphi$ 是显然的.

(2)令 $\alpha = \inf\phi = \inf\varphi$, $\beta = \sup\phi = \sup\varphi$,则有 α 是使得 $A - \alpha I$ 为半正定阵的最大实数,β 是使得 $\beta I - A$ 为半正定阵的最小实数.一方面,$|\langle Ax, x\rangle| \leqslant \| Ax \|_2 \ \| x \|_2 \leqslant \| A \|_2$

$\|x\|_2^2$，因此由 α,β 的定义立即知 $|\alpha|,|\beta|\leqslant\|A\|_2$；另一方面，若记 $r=\max\{|\alpha|,|\beta|\}$，则有 $rI+A$ 是半正定阵，$rI-A$ 也是半正定阵．注意到 $rI+A$ 与 $rI-A$ 可交换，因此 $(rI+A)(rI-A)$ 也是半正定阵，结果 r^2I-A^2 也是半正定阵，所以

$$r^2\|x\|_2^2=r^2\langle x,x\rangle\geqslant\langle A^2x,x\rangle=\langle Ax,Ax\rangle=\|Ax\|_2^2$$

$$r^2\geqslant\|A\|_2^2,\ r\geqslant\|A\|_2$$

定理获证．

□

作为谱范数的一个应用，下面说明矩阵的谱范数可用来给出向量空间中等维子空间距离的一种度量．首先证明如下定理．

定理 4.2.4　设 $U_1,U_2\in\mathbb{F}^{n\times r}$，且 $U_1^*U_1=U_2^*U_2=I_r$，则

$$\|U_1U_1^*-U_2U_2^*\|_2=\sqrt{1-\sigma_{\min}^2(U_1^*U_2)}$$

其中 $\sigma_{\min}^2(U_1^*U_2)$ 表示矩阵 $U_1^*U_2$ 的最小奇异值的平方．

证明：将 U_1,U_2 分别扩充为酉阵 $W=(V_1,U_1),Z=(V_2,U_2)$．考虑酉阵

$$P=\begin{pmatrix}P_{11}&P_{12}\\P_{21}&P_{22}\end{pmatrix}\triangleq\begin{pmatrix}V_1^*V_2&V_1^*U_2\\U_1^*V_2&U_1^*U_2\end{pmatrix}=\begin{pmatrix}V_1^*\\U_1^*\end{pmatrix}(V_2\quad U_2)$$

由

$$(P_{12}^*\quad P_{22}^*)\begin{pmatrix}P_{12}\\P_{22}\end{pmatrix}=I_r$$

得

$$P_{22}^*P_{22}=I_r-P_{12}^*P_{12}$$

因此，$P_{22}^*P_{22}$ 的特征值为不大于 1 的非负实数．取 P_{22} 的奇异值分解

$$T_1^*P_{22}H_1=C=\begin{pmatrix}c_1&&\\&\ddots&\\&&c_r\end{pmatrix}$$

其中 $0\leqslant c_1\leqslant\cdots\leqslant c_k\leqslant c_{k+1}=\cdots=c_r=1$．注意到

$$\begin{aligned}I_r&=H_1^*(P_{12}^*\quad P_{22}^*)\begin{pmatrix}P_{12}\\P_{22}\end{pmatrix}H_1\\&=H_1^*P_{12}^*P_{12}H_1+H_1^*P_{22}^*P_{22}H_1\\&=(P_{12}H_1)^*P_{12}H_1+C^2\end{aligned}$$

故有

$$(P_{12}H_1)^*P_{12}H_1=I_r-C^2=\begin{pmatrix}1-c_1^2&&&&\\&\ddots&&&\\&&1-c_k^2&&\\&&&0&\\&&&&\ddots&\\&&&&&0\end{pmatrix}$$

这表明：$P_{12}H_1$ 的前 k 列正交，后 $r-k$ 列均为零。现将 $P_{12}H_1$ 的前 k 列标准单位化后扩充为 $\mathbb{C}^{n-r}$ 的标准正交基 T_2，则有

$$T_2^* P_{12} H_1 = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}_{(n-r)\times r}$$

其中 $S = \begin{pmatrix} \sqrt{1-c_1^2} & & \\ & \ddots & \\ & & \sqrt{1-c_k^2} \end{pmatrix}_{k\times k}$，因此

$$\| V_1^* U_2 \|_2 = \| P_{12} \|_2 = \left\| \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \right\|_2 = \sqrt{1-c_1^2} = \sqrt{1-\sigma_{\min}^2(U_1^* U_2)}$$

现考虑 P 的行 $(P_{21} \quad P_{22})$，由

$$\begin{aligned} I_r &= T_1^* (P_{21} \quad P_{22}) \begin{pmatrix} P_{21}^* \\ P_{22}^* \end{pmatrix} T_1 \\ &= T_1^* P_{21} P_{21}^* T_1 + T_1^* P_{22} P_{22}^* T_1 \\ &= (T_1^* P_{21})(T_1^* P_{21})^* + C^2 \end{aligned}$$

有

$$(T_1^* P_{21})(T_1^* P_{21})^* = I_r - C^2 = \begin{pmatrix} 1-c_1^2 & & & & & \\ & \ddots & & & & \\ & & 1-c_k^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

因此，$T_1^* P_{21}$ 的前 k 行相互正交，后 $r-k$ 行均为零。将该 k 行分别标准单位化（即分别除以 $\dfrac{1}{\sqrt{1-c_i^2}}$），再将之扩充为行向量空间 $\mathbb{C}^{n-r}$ 的标准正交基，然后又将此标准正交基通过转置共轭得酉阵 H_2，则 H_2 满足

$$T_1^* P_{21} H_2 = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix}_{r\times(n-r)}$$

其中 $S = \begin{pmatrix} \sqrt{1-c_1^2} & & \\ & \ddots & \\ & & \sqrt{1-c_k^2} \end{pmatrix}_{k\times k}$.

于是

$$\| U_1^* V_2 \|_2 = \| P_{21} \|_2 = \left\| \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \right\|_2 = \sqrt{1-c_1^2} = \sqrt{1-\sigma_{\min}^2(U_1^* U_2)}$$

结果

$$\begin{aligned}\| U_1U_1^* - U_2U_2^* \|_2 &= \left\| \begin{pmatrix} V_1^* \\ U_1^* \end{pmatrix} (U_1U_1^* - U_2U_2^*)(V_2 \quad U_2) \right\|_2 \\ &= \left\| \begin{pmatrix} 0 & -V_1^* U_2 \\ U_1^* V_2 & 0 \end{pmatrix} \right\|_2 \\ &= \sqrt{1-\sigma_{\min}^2(U_1^* U_2)}\end{aligned}$$

定理证毕.

□

在向量空间中,除了向量间的夹角及向量间的距离,还可引进等维子空间的夹角和等维子空间的距离等概念.

定义 4.2.2　设 L,M 是 $\mathbb{F}^n$ 的两个等维子空间,E_1,E_2 分别是 $\mathbb{F}^n$ 某一标准正交基下正交投影算子 P_L、P_M 的表示矩阵,U_1,U_2 分别为 $R(E_1)$ 与 $R(E_2)$ 的标准正交基构成的矩阵.称

$$\theta = \arccos \sigma_{\min}(U_1^* U_2)$$

为等维子空间 L 与 M 之间的**夹角**,其中 $\sigma_{\min}(U_1^* U_2)$ 为 $U_1^* U_2$ 的最小奇异值,$R(E_i)$ 表示由 E_i 的列所张成的向量空间.

对任何句逆阵 $T \in \mathbb{F}^{n\times n}$,由于 $R(E_iT) = R(E_i)$,因此容易知道,θ 与 $\mathbb{F}^n$ 的标准正交基的选取方式无关,也与 $R(E_1)$ 和 $R(E_2)$ 的标准正交基的选取无关.

定义 4.2.3　设 L,M 是 $\mathbb{F}^n$ 的两个子空间,$\dim L = \dim M$,E_1,E_2 分别是 $\mathbb{F}^n$ 某一标准正交基下正交投影算子 P_L,P_M 的表示矩阵,称

$$d(L,M) = \| E_1 - E_2 \|_2$$

为等维子空间 L 与 M 的**距离**.

由于 $\mathbb{F}^n$ 的不同标准正交基下同一正交投影的矩阵表示是酉相似的.因此,由定理 4.2.2 中(4)知,上述定义的距离与标准正交基的选取方式无关.

现给定 H 为一个 Hermite 幂等阵,由分解定理知,H 可表示为 $H = V\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} V^*$,其中 V 为酉阵.令 $V = (v_1,\cdots,v_n)$,则有 $H = UU^*$,其中 $U = (v_1,\cdots,v_r)$,满足 $U^*U = I_r$.因此,等维子空间 L 与 M 的距离 $d(L,M)(= \| E_1 - E_2 \|_2)$ 在将 E_1,E_2 分别分解为 $E_1 = U_1U_1^*$ 和 $E_2 = U_2U_2^*$(U_1 与 U_2 均为部分酉阵)后,变为 $\| U_1U_1^* - U_2U_2^* \|_2$.再由定理 4.2.4 得:等维子空间 L 与 M 的距离实际上就是这两个子空间夹角的正弦.

4.3　相容矩阵范数

由于任何 $\mathbb{F}$ 上 $m\times n$ 矩阵既可看成是域 $\mathbb{F}$ 上 mn 维线性空间中的一个向量,又可看成 $\mathbb{F}^n \to \mathbb{F}^m$ 在给定基下的一个线性变换.因此,人们往往关注能反映这二重特征的矩阵范数,为此引入如下矩阵范数的相容性概念.

定义 4.3.1 设 $\|\cdot\|_{m\times n}$, $\|\cdot\|_{m\times p}$和$\|\cdot\|_{p\times n}$分别是$\mathbb{F}^{m\times n}$, $\mathbb{F}^{m\times p}$和$\mathbb{F}^{p\times n}$中的给定范数.若$\forall A\in\mathbb{F}^{m\times p}$, $\forall B\in\mathbb{F}^{p\times n}$,均有

$$\|AB\|_{m\times n}\leqslant\|A\|_{m\times p}\|B\|_{p\times n}$$

则称$\|\cdot\|_{m\times n}$、$\|\cdot\|_{m\times p}$和$\|\cdot\|_{p\times n}$是**相容矩阵范数**.对$\mathbb{F}^{n\times n}$中的范数$\|\cdot\|$,若满足$\|AB\|\leqslant\|A\|\|B\|$对任何$A,B\in\mathbb{F}^{n\times n}$皆成立,则称$\|\cdot\|$是**自相容的矩阵范数**.今后本书中涉及的相容矩阵范数,除非特别说明,均是指自相容的矩阵范数.

显然,上述关于相容矩阵范数的定义包含了下述情形:

设$\|\cdot\|_{m\times n}$为$\mathbb{F}^{m\times n}$上的范数,$\|\cdot\|_\alpha$与$\|\cdot\|_\beta$分别是$\mathbb{F}^n$与$\mathbb{F}^m$上的向量范数.若$\|Ax\|_\beta\leqslant\|A\|_{m\times n}\|x\|_\alpha$对任何的$A\in\mathbb{F}^{m\times n}$及任何的$x\in\mathbb{F}^n$皆成立,则有$\|\cdot\|_\beta$、$\|\cdot\|_{m\times n}$和$\|\cdot\|_\alpha$是相容的.特别当$A\in\mathbb{F}^{n\times n}$,且$\|\cdot\|_\alpha=\|\cdot\|_\beta$时,上述不等式变为

$$\|Ax\|_\alpha\leqslant\|A\|_{n\times n}\|x\|_\alpha$$

此时称方阵A的范数$\|A\|_{n\times n}$与向量的范数$\|x\|_\alpha$相容.$\|\cdot\|_\alpha$也称为与$\|\cdot\|_{n\times n}$相容的向量范数.

例4.3.1 设$A\in\mathbb{C}^{m\times n}$, $x\in\mathbb{C}^n$, $\|A\|$为A的算子范数.由算子范数定义知:$\|Ax\|_\beta\leqslant\|A\|\|x\|_\alpha$对任何$x$皆成立.因此,$A$的算子范数与向量的$\beta$-范数和$\alpha$-范数是相容的.此外,仍由算子范数的定义知算子范数为相容的矩阵范数.即只要AB有定义,就有$\|AB\|\leqslant\|A\|\|B\|$.

例 4.3.2 $\mathbb{F}^{m\times n}$中,矩阵的 Frobenius 范数$\|A\|_F=\left(\sum\limits_{i,j}|a_{ij}|^2\right)^{\frac{1}{2}}$是一个与向量的 Euclid 范数相容的范数,矩阵的 Frobenius 范数也是自相容的矩阵范数.事实上,令$x=(x_1,\cdots,x_n)^{\mathrm{T}}\in\mathbb{F}^n$,

$$\begin{aligned}\|Ax\|_2^2&=\sum_i\Big|\sum_j a_{ij}x_j\Big|^2\leqslant\sum_i\Big(\sum_j|a_{ij}||x_j|\Big)^2\\&\leqslant\sum_i\Big[\Big(\sum_j|a_{ij}|^2\Big)^{\frac{1}{2}}\Big(\sum_j|x_j|^2\Big)^{\frac{1}{2}}\Big]^2\\&=\sum_i\Big[\Big(\sum_j|a_{ij}|^2\Big)\Big(\sum_j|x_j|^2\Big)\Big]\\&=\Big(\sum_j|x_j|^2\Big)\sum_{i,j}|a_{ij}|^2\end{aligned}$$

结果

$$\|Ax\|_2\leqslant\|A\|_F\|x\|_2$$

进一步有

$$\|AB\|_F\leqslant\|A\|_F\|B\|_F$$

□

将矩阵经向量的 Euclid 范数所导出的算子范数$\|A\|_2$(i.e. 谱范数)与矩阵的 Frobenius范数$\|A\|_F$相比较,易知

$$\sqrt{\lambda_1}=\|A\|_2\leqslant\|A\|_F$$

其中λ_1为A^*A的最大特征值.一般地,矩阵A的 Frobenius 范数与谱范数不等.例如,单

位矩阵 I_n，其 Frobenius 范数 $\| I_n \|_F$ 为 $\sqrt{n}$，而其谱范数 $\| I_n \|_2$ 则为 1.

例 4.3.3　在 $\mathbb{F}^{2\times 2}$ 中，$A=(a_{ij})\in \mathbb{F}^{2\times 2}$. 取 $\| A \|_{v_\infty} = \max\limits_{i,j} |a_{ij}|$. 于是 $\|\cdot\|_{v_\infty}$ 为 $\mathbb{F}^{2\times 2}$ 上的矩阵范数，但 $\|\cdot\|_{v_\infty}$ 不是相容的矩阵范数. 如取

$$A=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\ B=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\ AB=\begin{pmatrix}2 & 2\\ 2 & 2\end{pmatrix}$$

容易看出 $\| A \|_{v_\infty} = \| B \|_{v_\infty} = 1$，而 $\| AB \|_{v_\infty} = 2$.

此例说明，并非所有的矩阵范数均是相容的矩阵范数. 不过，目前我们所熟悉的，诸如矩阵的列范数、行范数等均是相容的.

定理 4.3.1　矩阵范数 $\|\cdot\|_1$（列范数），$\|\cdot\|_\infty$（行范数）以及 $\|\cdot\|_{v_1}$ 均为相容的矩阵范数.

证明: (1) 由于矩阵列范数 $\|\cdot\|_1$ 和矩阵行范数 $\|\cdot\|_\infty$ 均可看成算子子范数，因此 $\|\cdot\|_1$ 是相容的矩阵范数，$\|\cdot\|_\infty$ 也是相容的矩阵范数.

(2) $A = a_{ij}, B = b_{ij}, AB = c_{ij}, c_{ij} = \sum\limits_{k=1}^{l} a_{ik}b_{kj}$，于是有

$$\begin{aligned}\| AB \|_{v_1} &= \sum_{i,j} \Big| \sum_k a_{ik}b_{kj} \Big| \\ &\leqslant \sum_{i,j,k} |a_{ik}|\,|b_{kj}| \\ &= \sum_{i,k}\sum_j |a_{ik}|\,|b_{kj}| \\ &= \sum_{i,k}\Big[\,|a_{ik}|\Big(\sum_j |b_{kj}|\Big)\Big] \\ &\leqslant \Big(\sum_{i,k}|a_{ik}|\Big)\| B \|_{v_1} \\ &= \| A \|_{v_1}\| B \|_{v_1}\end{aligned}$$

□

相容矩阵范数还具有如下性质.

性质 4.3.1　设 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 上的相容矩阵范数，则在 $\mathbb{F}^n$ 上必存在与之相容的向量范数，即存在向量范数 $\|\cdot\|_a$，使得 $\| Ax \|_a \leqslant \| A \|\,\| x \|_a$ 对 $\forall A\in \mathbb{F}^{n\times n}$ 及 $\forall x\in \mathbb{F}^n$ 皆成立.

证明: 取 $\mathbb{F}^n$ 中向量 $y=(1\cdots 1)^{\mathrm{T}}$，则由 $\|\cdot\|$ 是 $\mathbb{F}^{n\times n}$ 上给定的范数知：$\forall x\in \mathbb{F}^n$，若令 $\| x \|_y = \| xy^{\mathrm{T}} \|$，则有 $\|\cdot\|_y$ 为 $\mathbb{F}^n$ 上的一个向量范数，对此向量范数有

$$\| Ax \|_y = \| Ax\cdot y^{\mathrm{T}} \| \leqslant \| A \|\,\| xy^{\mathrm{T}} \| = \| A \|\,\| x \|_y$$

此即 $\|\cdot\|_y$ 与 $\|\cdot\|$ 相容.

□

性质 4.3.2　设 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$（$\mathbb{F}=\mathbb{R}$ 或 $\mathbb{C}$）上的相容矩阵范数，则 $\forall A\in \mathbb{F}^{n\times n}$ 有 $\rho(A) \leqslant \| A \|$，其中 $\rho(A)$ 表示 A 的谱半径.

证明: (1) 当 $\mathbb{F}=\mathbb{C}$ 时，由性质 4.3.1 知，存在 $\mathbb{C}^n$ 上范数 $\|\cdot\|_v$ 使得 $\| Ax \|_v \leqslant$

$\|A\| \|x\|_v$对任何$x\in\mathbb{C}^n$皆成立.对A的任一特征值λ,取相应于λ的一个非零特征向量,有$A\alpha=\lambda\alpha$,此时

$$\|A\alpha\|_v=|\lambda| \|\alpha\|_v\leqslant\|A\| \|\alpha\|_v$$

由$\|\alpha\|_v\neq0$立即知$|\lambda|\leqslant\|A\|$.获证.

(2)当数域$\mathbb{F}=\mathbb{R}$时,于是由$\mathbb{F}^{n\times n}$上的范数等价定理知,存在$M>0$使得$\|A^k\|\geqslant M\|A^k\|_2$.注意到由(1)知$\|A^k\|_2\geqslant\rho(A^k)$,于是

$$\|A\|^k\geqslant M\|A^k\|_2\geqslant M\rho(A^k)=M(\rho(A))^k$$

结果有$\|A\|\geqslant M^{\frac{1}{k}}\rho(A)$,让$k\to\infty$得$\|A\|\geqslant\rho(A)$.

□

定理 4.3.2 $\forall A\in\mathbb{F}^{n\times n},\forall\varepsilon>0$,则存在$\mathbb{F}^{n\times n}$上的相容矩阵范数$\|\cdot\|$(它依赖于矩阵$A$和$\varepsilon$)满足$\|I_n\|=1$,使得$\|A\|\leqslant\rho(A)+\varepsilon$.

证明: 由 Schur 分解定理知,有酉阵$U\in\mathbb{C}^{n\times n}$,使得

$$U^*AU=R=\begin{pmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \vdots \\ & & r_{nn}\end{pmatrix}$$

令

$$D\triangleq\begin{pmatrix}1 & & & \\ & \delta & & \\ & & \ddots & \\ & & & \delta^{n-1}\end{pmatrix}(\delta\neq0)$$

则有

$$D^{-1}RD=\begin{pmatrix} r_{11} & r_{12}\delta & & \cdots & r_{1n}\delta^{n-1} \\ & r_{22} & & \cdots & r_{2n}\delta^{n-2} \\ & & \ddots & & \vdots \\ & & & r_{(n-1)(n-1)} & r_{(n-1)n}\delta \\ & & & & r_{nn}\end{pmatrix}$$

现取δ为充分小的正数使得$\|D^{-1}RD\|_1\leqslant\max\limits_{1\leqslant i\leqslant n}\{|r_{ii}|+\varepsilon\}$,结果$\|D^{-1}U^*AUD\|_1=\|D^{-1}RD\|_1\leqslant\rho(A)+\varepsilon$.

$\forall G\in\mathbb{F}^{n\times n}$,再记$\|G\|=\|D^{-1}U^*GUD\|_1$,则易知$\|\cdot\|$为$\mathbb{F}^{n\times n}$上的相容矩阵范数,且满足$\|I_n\|=1$,结果$\|A\|\leqslant\rho(A)+\varepsilon$.证毕.

□

性质 4.3.2 和定理 4.3.2 初步阐明了矩阵谱半径与矩阵相容范数之间的关系:

$$\forall A\in\mathbb{F}^{n\times n},\rho(A)=\inf\{\|A\|:\|\cdot\|为\mathbb{F}^{n\times n}上的相容矩阵范数\}$$

4.4 矩阵极限与矩阵级数

熟知,对一个赋范线性空间而言,其向量序列的极限是依范数 $\|\cdot\|$ 而给出的,即向量序列 $\{x_n\}$ 收敛于 x 是指 $\lim\limits_{n\to\infty}\|x_n - x\| = 0$. 由于矩阵范数实际上为向量范数,因此矩阵序列的极限便可类似地给出.

定义 4.4.1 设 $\|\cdot\|$ 为 $\mathbb{F}^{m\times n}$ 中的矩阵范数. $\mathbb{F}^{m\times n}$ 中的矩阵序列 $\{A_i\}_{i=1}^{\infty}$ 若满足:存在 $A\in\mathbb{F}^{m\times n}$ 使得 $\lim\limits_{i\to\infty}\|A_i - A\| = 0$, 则称矩阵序列 $\{A_i\}_{i=1}^{\infty}$ **按范数** $\|\cdot\|$ **收敛于** A, 记作 $A_i \xrightarrow{\|\cdot\|} A$. 不收敛的矩阵序列称为是**发散的**.

上述定义表明,赋范空间的向量收敛现也可看成是矩阵序列收敛的特殊情形.

关于矩阵序列收敛性的判定,有如下定理:

定理 4.4.1 (1)在有限维向量空间中,向量序列 $\{x_m\}$ 按某种范数收敛于 x 等价于向量序列按任何一种范数收敛于 x. 特别地,等价于 $\{x_m\}$ 按坐标收敛于 x.

(2) $\mathbb{F}^{m\times n}$ 中矩阵序列 $\{A_m = a_{ij}^{(m)}\}$ 按某种范数收敛于 $A = (a_{ij})$ 等价于 $\{A_m\}$ 按任何一种矩阵范数收敛于 A. 特别地,等价于 $\{A_m\}$ 依每个元素收敛于 A, i.e. $\lim\limits_{m\to\infty} a_{ij}^{(m)} = a_{ij}$ 对 $\forall i, j$ 皆成立.

证明: (1)由有限维向量空间的范数等价,特别的情况下,又由按向量的 Euclid 范数收敛等价于按坐标收敛,因此(1)真.

(2)是(1)的直接结果.

□

关于矩阵序列的极限还有如下性质:

性质 4.4.1 设 $\{A_j\}$ 与 $\{B_j\}$ 为 $\mathbb{F}^{m\times n}$ 中两个矩阵序列. 若 $\lim A_j = A$, $\lim B_j = B$, 则有

(1) $\lim(kA_j + lB_j) = kA + lB$;

(2) $\lim(A_jB_j) = AB$, 当 A_j, B_j 均为同阶方阵时.

证明: 显然.

□

性质 4.4.2 在 $\mathbb{F}^{m\times n}$ 中,设 $\lim A_i = A$, 且 A^{-1}, A_i^{-1} 均存在, $i = 1, 2, \cdots$, 则有 $\lim A_i^{-1}$ 也存在且为 A^{-1}.

证明: 注意到

$$A_i^{-1} = \frac{1}{\det A_i} A_i^{adj}$$

其中 A_i^{adj} 表示 A_i 的伴随矩阵. 由行列式函数的连续性知, $\lim\limits_i \det A_i = \det A$ 及 $\lim\limits_i A_i^{adj} = A^{adj}$, 因此, $\lim\limits_i A_i^{-1} = \frac{1}{\det A} A^{adj} = A^{-1}$.

□

性质 4.4.3 设 $({}_{\mathbb{F}}V, \langle\cdot\rangle)$ 为有限维内积空间, $\{x_n\}$ 为 ${}_{\mathbb{F}}V$ 中的向量序列, $x\in V$; $\{A_i\}$ 为

$\mathbb{F}^{m\times n}$中的矩阵序列，$A\in\mathbb{F}^{m\times n}$，则有

(1)$\lim\limits_{n} x_n = x \Leftrightarrow \lim\limits_{n\to\infty}\langle x_n - x, y\rangle = 0$ 对任意的 $y\in V$;

(2)$\lim\limits_{i} A_i = A \Leftrightarrow \lim\limits_{i} \operatorname{tr}(A_i - A)B^* = 0$ 对任意的 $B\in\mathbb{F}^{m\times n}$.

证明: (1)由关于内积的许瓦兹不等式立即便知.

(2)注意$\langle X, Y\rangle = \operatorname{tr}(XY^*)$是$\mathbb{F}^{m\times n}$的内积，因此由(1)即知.

□

定理 4.4.2 设 U 是 n 阶酉阵，M 为 $Ux = x$ 的解空间，则由

$$D_n = \frac{1}{n}(I + U + \cdots + U^{n-1})$$

定义的矩阵序列$\{D_n\}_{n=1}^{\infty}$收敛于垂直射影 $E = P_M$，其中 E 为$\mathbb{C}^n$在$\{e_1, \cdots, e_n\}$下垂直射影算子 P_M 的表示阵，$e_i = (0, \cdots, \underset{i}{1}, 0, \cdots, 0)^{\mathrm{T}}$.

证明: 记 $N = \operatorname{Im}(I - U)$. 若 $x = y - Uy \in N$，则

$$D_n x = \frac{1}{n}(y - Uy + Uy - U^2 y + \cdots + U^{n-1} y - U^n y) = \frac{1}{n}(y - U^n y)$$

该向量的 Euclid 范数为

$$\| D_n x \|_2 = \frac{1}{n} \| y - U^n y \|_2 \leqslant \frac{1}{n}(\| y \|_2 + \| U^n y \|_2) = \frac{2}{n} \| y \|_2$$

因此，对 $\forall x\in N$ 有 $D_n x \to 0$.

另一方面，若 $x\in M$，则 $Ux = x$，从而 $D_n x = x$，$D_n x \to x$. 注意到 U 是酉阵，因此又有 $M \perp N$，$N \oplus M = \mathbb{C}^n$. 现在记 $E = P_M$ 为 $\mathbb{C}^n$ 到 M 上关于基$\{e_1, \cdots, e_n\}$的垂直射影矩阵，于是$\forall x\in\mathbb{C}^n$，$x = Ex \oplus (I - E)x$，其中 $Ex\in M$，$(I - E)x\in N$，$\| D_n x - Ex \|_2 = \| D_n Ex + D_n(I - E)x - Ex \|_2 \to 0$，亦即$\| (D_n - E)x \|_2 \to 0$. 由 x 在$\mathbb{C}^n$中的任意性知 $D_n \to E$.

□

直观地，给定$\mathbb{F}^{m\times n}$中的一个矩阵序列 $A_1, A_2, \cdots, A_k, \cdots$，称表达式 $A_1 + A_2 + \cdots + A_k + \cdots$为一个**矩阵级数**，记为$\sum\limits_{k=1}^{\infty} A_k$，其中 A_k 叫做该级数的**通项**或**一般项**. 容易理解，矩阵级数的收敛性概念也是通过矩阵序列的极限来给出的.

定义 4.4.2 对一个矩阵级数 $\sum\limits_{i=1}^{\infty} A_i$，若其部分和序列$\{S_k = A_1 + A_2 + \cdots + A_k\}$极限存在，记为 S，则称该**矩阵级数收敛**，记作$\sum\limits_{i=1}^{\infty} A_i = S$. 否则称级数$\sum\limits_{i=1}^{\infty} A_i$ 是**发散的**.

关于矩阵级数收敛性的判定有如下结果.

定理 4.4.3 (1)(Cauchy 收敛准则) $\sum\limits_{i=1}^{\infty} A_i$ 收敛 $\Leftrightarrow$ 对$\mathbb{F}^{m\times n}$ 的任何一种范数 $\|\cdot\|$ 及 $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ s.t. 对任何正整数 p, q，当 $q \geqslant p \geqslant N$ 时，均有$\left\| \sum\limits_{i=p}^{q} A_i \right\| < \varepsilon \Leftrightarrow$ 对某一种范数 $\|\cdot\|$ 满足：$\forall \varepsilon > 0, \exists N \in \mathbb{N}$，当 $q \geqslant p \geqslant N$ 时有$\left\| \sum\limits_{i=p}^{q} A_i \right\| < \varepsilon$，这里 p, q 皆为

正整数.

(2) 若数项级数 $\sum_{i=1}^{\infty} \| A_i \|$ 收敛,则矩阵级数 $\sum_{i=1}^{\infty} A_i$ 收敛.

(3)设数项幂级数 $\sum_{k=1}^{\infty} a_k x^k$ 的收敛半径为 R,则当方阵 A 的某相容矩阵范数 $\| A \| < R$ 时, $\sum_{k=1}^{\infty} a_k A^k$ 收敛.

证明: (1)考虑 $\sum_{i=1}^{\infty} A_i$ 的部分和序列 $\{S_m\}$, $S_m = A_1 + \cdots + A_m$.

$\sum_{i=1}^{\infty} A_i$ 收敛 $\Leftrightarrow$ $\{S_m\}$ 极限存在;

$\Leftrightarrow$ $\{S_m\}$ 按任何一种范数收敛;

$\Leftrightarrow$ $\{S_m\}$ 按某种范数 $\| \cdot \|$ 收敛;

$\Leftrightarrow$ $\forall \varepsilon > 0, \exists N \in \mathbb{N}$,只要 $m_1, m_2 > N$ 就有 $\| S_{m_1} - S_{m_2} \| < \varepsilon$;

$\Leftrightarrow$ $\forall \varepsilon > 0, \exists N \in \mathbb{N}$ 使得对任何正整数 p, q,当 $q \geqslant p \geqslant N$ 时,就有

$$\left\| \sum_{i=p}^{q} A_i \right\| < \varepsilon$$

(2)由(1)立即可知.

(3)当 $\| A \| < R$ 时, $\sum_{k=1}^{\infty} a_k \| A \|^k$ 收敛.注意到此收敛还为绝对收敛,由(2)即得(3).

□

由矩阵级数的 Cauchy 收敛准则知:

推论 4.4.1　设 $\sum_{k=1}^{\infty} A_k$ 为 $\mathbb{F}^{m\times n}$ 中的矩阵级数.若 $\sum_{k=1}^{\infty} A_k$ 收敛,则必有 $\lim A_k = 0$.

在矩阵级数中,也可引入绝对收敛的概念.对一个给定的矩阵级数 $\sum_{k=1}^{\infty} A_k$,其中 $A_k = (a_{ij}^{(k)})_{m\times n}$,通常称该级数是**绝对收敛**的是指: $\sum_{k=1}^{\infty} a_{ij}^{(k)}$ 对任意的 (i,j), $i = 1,2,\cdots,m; j = 1,2,\cdots,n$ 均是绝对收敛的.

由数项级数绝对收敛的性质容易知道.

性质 4.4.4　若矩阵级数 $\sum_{k=1}^{\infty} A_k$ 绝对收敛,则(1) $\sum_{k=1}^{\infty} A_k$ 收敛;(2) $\sum_{k=1}^{\infty} A_k$ 任意重排后得到的矩阵级数仍收敛,且其和不变.

性质 4.4.5　设 $\sum_{k=1}^{\infty} A_k$ 为 $\mathbb{F}^{m\times n}$ 中的矩阵级数, $\| \cdot \|$ 为 $\mathbb{F}^{m\times n}$ 中的矩阵范数,则 $\sum_{k=1}^{\infty} A_k$ 绝对收敛的充要条件是 $\sum_{k=1}^{\infty} \| A_k \|$ 收敛.

证明: $\Rightarrow$ $\sum_{k=1}^{\infty} A_k$ 绝对收敛 $\Leftrightarrow \forall\, i,j, \sum_{k=1}^{\infty} |a_{ij}^{(k)}|$ 收敛. 因此, $\forall\, \varepsilon > 0, \exists\, k_0 \in \mathbb{Z}^+$ 使得 $p > k_0$ 时有 $\sum_{k=p}^{p+q}\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}^{(k)}|\right) < \varepsilon$. 进而

$$\sum_{k=p}^{p+q} \| A_k \|_{\infty} = \sum_{k=p}^{p+q} \max_{1\leqslant i\leqslant m}\left(\sum_{j=1}^{n}|a_{ij}^{(k)}|\right) \leqslant \sum_{k=p}^{p+q}\left(\sum_{i,j}|a_{ij}^{(k)}|\right) < \varepsilon$$

此即 $\sum_{k=1}^{\infty} \| A_k \|_{\infty}$ 收敛, 再由矩阵的任何两种范数均是等价地推导出: $\sum_{k=1}^{\infty} \| A_k \|$ 收敛.

$\Leftarrow$ $\sum_{k=1}^{\infty} \| A_k \|$ 收敛等价于 $\sum_{k=1}^{\infty} \| A_k \|_{\infty}$ 收敛, 从而

$$\sum_{k=p}^{p+q} |a_{ij}^{(k)}| \leqslant \sum_{k=p}^{p+q} \| A_k \|_{\infty}$$

这表明 $\sum_{k=1}^{\infty} A_k$ 为绝对收敛的.

□

性质 4.4.6 设 $\sum_{k=1}^{\infty} A_k$ 为 $\mathbb{F}^{m\times l}$ 中绝对收敛的级数, $A = \sum_{k=1}^{\infty} A_k$, $\sum_{k=1}^{\infty} B_k$ 是 $\mathbb{F}^{l\times n}$ 中绝对收敛的级数, $B = \sum_{k=1}^{\infty} B_k$. 则级数乘积 $\left(\sum_{k=1}^{\infty} A_k\right)\left(\sum_{k=1}^{\infty} B_k\right)$ 按任何方式排列后得到的级数也绝对收敛, 且其和仍为 AB.

证明: 利用性质 4.4.5 知, $\sum_{k=1}^{\infty} \| A_k \|_{\infty}$ 和 $\sum_{k=1}^{\infty} \| B_k \|_{\infty}$ 均收敛. 让 $M = \sum_{k=1}^{\infty} \| A_k \|_{\infty}$, $N = \sum_{k=1}^{\infty} \| B_k \|_{\infty}$, 级数乘积 $\left(\sum_{k=1}^{\infty} A_k\right)\left(\sum_{k=1}^{\infty} B_k\right)$ 的任一排列方式得到的级数记为 $\sum_{k=1}^{\infty} C_k$, 于是由 $\| \cdot \|_{\infty}$ 是相容矩阵范数知

$$\begin{aligned}
\sum_{k=1}^{p} \| C_k \|_{\infty} &\leqslant \sum_{k=1}^{p}\sum_{i,j} \| A_i B_j \|_{\infty} \\
&\leqslant \sum_{k=1}^{p}\sum_{i,j} \| A_i \|_{\infty} \| B_j \|_{\infty} \\
&\leqslant \sum_{k=1}^{\infty} \| A_k \|_{\infty} \sum_{k=1}^{\infty} \| B_k \|_{\infty} \\
&\leqslant MN, \text{有界}
\end{aligned}$$

这表明数项级数 $\sum_{k=1}^{\infty} \| C_k \|_{\infty}$ 收敛. 再由性质 4.4.5 得 $\sum_{k=1}^{\infty} C_k$ 绝对收敛.

显然, $\left(\sum_{k=1}^{\infty} A_k\right)\left(\sum_{k=1}^{\infty} B_k\right)$ 按任何方式排列后得到的级数实质上是 $\sum_{k=1}^{\infty} C_k$ 经过适当重排后得到的级数. 因此, 由性质 4.4.4 之(2) 知它收敛于同一个矩阵. 若按如下顺序排列 $\left(\sum_{k=1}^{\infty} A_k\right)\left(\sum_{k=1}^{\infty} B_k\right)$:

$$\sum_{k=1}^{\infty} C_k = A_1 B_1 + (A_1 B_2 + A_2 B_1 + A_2 B_2) + \cdots$$

结果:

$$\lim_{p\to\infty}\left(\sum_{k=1}^{p} C_k\right) = \lim_{p\to\infty}\left(\sum_{k=1}^{p} A_k \sum_{k=1}^{p} B_k\right) = \left[\lim_{p\to\infty}\left(\sum_{k=1}^{p} A_k\right)\right]\left[\lim_{p\to\infty}\left(\sum_{k=1}^{p} B_k\right)\right] = AB$$

□

4.5　相容矩阵范数与矩阵序列极限的应用

此节,我们借用相容矩阵范数给出矩阵级数的进一步有关结果.作为矩阵序列的应用,我们证明非负矩阵论中著名的 Perron 定理.回顾上一节的性质 4.4.5,它将级数 $\sum_{k=1}^{\infty} A_k$ 的绝对收敛性判定转换成了数项级数 $\sum_{k=1}^{\infty} \| A_k \|$ 的收敛性判定.基于这一点,我们得到:

定理 4.5.1(Lagrange - Sylvester 定理)　设数项幂级数 $\sum_{k=1}^{\infty} a_k t^k$ 的收敛半径为 r, $A \in \mathbb{F}^{n\times n}$,则有当 $\rho(A) < r$ 时,矩阵级数 $\sum_{k=1}^{\infty} a_k A^k$ 绝对收敛.更一般地,设 $f(t-t_0) = \sum_{k=1}^{\infty} a_k (t-t_0)^k$ 的收敛半径为 r,则当 A 的所有特征值均落在收敛圆盘(i.e. $|\lambda - t_0| < r, \lambda \in \operatorname{spec}A$) 内时,有 $\sum_{k=1}^{\infty} a_k (A - t_0 I)^k$ 绝对收敛.进一步,假定 $A = PJP^{-1}$, $J = \bigoplus_{i=1} J_{n_i}(\lambda_i)$ 为 A 的 Jordan 标准形, $U_{n_i} = J_{n_i}(0)$, $f^{(j)}$ 表示 $f(t)$ 的 j 阶导数,则

$$\sum_{k=1}^{\infty} a_k (A - t_0 I)^k = P\left[\bigoplus_{i=1} \sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_i - t_0)}{j!} U_{n_i}^j\right] P^{-1}$$

证明: 由于 $\rho(A - t_0 I) < r$,因此可取某一相容的矩阵范数 $\|\cdot\|$,使得

$$\rho(A - t_0 I) \leqslant \| A - t_0 I \| < r$$

结果

$$\sum_{k=0}^{\infty} \| a_k (A - t_0 I)^k \| \leqslant \sum_{k=0}^{\infty} |a_k| \, \| A - t_0 I \|^k \text{ 收敛}$$

亦即

$$\sum_{k=0}^{\infty} a_k (A - t_0 I)^k \text{ 绝对收敛}$$

剩下的只需注意到

$$f(\lambda_i - t_0) = \sum_k a_k (\lambda_i - t_0)^k$$

$$\frac{f^{(j)}(\lambda_i - t_0)}{j!} = \sum_k a_k C_k^j (\lambda_i - t_0)^{k-j}$$

当 $l \geqslant n_i - 1$ 时,有

$$J_{n_i}^l(\lambda_i - t_0) = [(\lambda_i - t_0) I_{n_i} + U_{n_i}]^l$$

$$= \sum_{j=0}^{n_i-1} C_l^j(\lambda_i - t_0)^{l-j} U_{n_i}^j$$

$$= \begin{pmatrix} (\lambda_i - t_0)^l & C_l^1(\lambda_i - t_0)^{l-1} & \cdots & C_l^{n_i-1}(\lambda_i - t_0)^{l-n_i+1} \\ & & & \vdots \\ & \ddots & \ddots & C_l^1(\lambda_i - t_0)^{l-1} \\ & & & (\lambda_i - t_0)^l \end{pmatrix} \tag{4.5.1}$$

$$\sum_k a_k J_{n_i}^k(\lambda_i - t_0) = \sum_{j=0}^{n_i-1} \frac{f^{(j)}(\lambda_i - t_0)}{j!} U_{n_i}^j$$

$$= \begin{pmatrix} f(\lambda_i - t_0) & f'(\lambda_i - t_0) & \cdots & \frac{1}{(n_i-1)!} f^{(n_i-1)}(\lambda_i - t_0) \\ & & \ddots & \vdots \\ & \ddots & & f'(\lambda_i - t_0) \\ & & & f(\lambda_i - t_0) \end{pmatrix}$$

整个定理获证.

□

例 4.5.1 由 $\sin z = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 + \cdots + (-1)^k \frac{1}{(2k+1)!}z^{2k+1} + \cdots$

$$\cos z = 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 + \cdots + (-1)^k \frac{1}{(2k)!}z^{2k} + \cdots$$

$$e^z = 1 + z + \frac{1}{2!}z^2 + \cdots + \frac{1}{k!}z^k + \cdots$$

它们的收敛半径为 ∞,因此,任给 $A \in \mathbb{C}^{n \times n}$,以下三个矩阵级数:

$$A - \frac{1}{3!}A^3 + \frac{1}{5!}A^5 + \cdots + (-1)^k \frac{1}{(2k+1)!}A^{2k+1} + \cdots$$

$$I - \frac{1}{2!}A^2 + \frac{1}{4!}A^4 + \cdots + (-1)^k \frac{1}{(2k)!}A^{2k} + \cdots$$

$$I + A + \frac{1}{2!}A^2 + \cdots + \frac{1}{k!}A^k + \cdots$$

均是绝对收敛的.通常记它们的和矩阵分别为 $\sin A, \cos A, e^A$.

如果数项级数 $\sum_k a_k z^k$ 的收敛半径为 R,则对任意的 $|z| < R$, $\sum_k a_k z^k$ 都是收敛的,这样便定义了一个函数:

$$f(z) = \sum_k a_k z^k, \quad |z| < R$$

同样,当取 $\mathbb{C}^{n \times n}$ 的子集 $D(f) = \{A \in \mathbb{F}^{n \times n} \mid \rho(A) < R\}$,则对每个 $B \in D(f)$,矩阵的幂级数 $\sum_k a_k B^k$ 也是收敛的.这样映射:

$$f: D(f) \to \mathbb{C}^{n \times n}$$
$$A \to f(A) = \sum_k a_k A^k$$

便定义了一个以矩阵 A 为变量而取值也为矩阵的函数.一般地,我们把 $f(A)$ 称作标量函

数的矩阵推广或矩阵函数.

由(4.5.1)式,下列命题的证明是容易的.

命题 4.5.1　设 $A\in\mathbb{C}^{n\times n}$,则 $\lim\limits_{k\to\infty}A^k$ 存在当且仅当下列情形之一发生:(1) $\rho(A)<1$;(2) $\rho(A)=1$,但落在单位圆周上的特征值只有1,并且对应于1的初等因子的次数都是1.

证明: 设 $A=PJP^{-1}$, $J=\bigoplus\limits_{i=1}^{k}J_{n_i}(\lambda_i)$ 为 A 的 Jordan 标准形.于是 $\lim\limits_{k\to\infty}A^k$ 存在⇔对每个 Jordan 块 $J_{n_i}(\lambda_i)$, $\lim\limits_{k\to\infty}J_{n_i}^k(\lambda_i)$ 存在.由(4.5.1)式知,当 $k\geqslant n_i$ 时,有

$$J_{n_i}^k(\lambda_i)=\begin{pmatrix}\lambda_i^k & C_k^1\lambda_i^{k-1} & \cdots & C_k^{n_i-1}\lambda_i^{k-n_i+1}\\ & \ddots & \ddots & \vdots\\ & & \ddots & C_k^1\lambda_i^{k-1}\\ & & & \lambda_i^k\end{pmatrix}$$

由此,要使 $\lim\limits_{k\to\infty}J_{n_i}^k(\lambda_i)$ 存在,要么 $|\lambda_i|<1$,要么满足 $\begin{cases}\lambda_i=1\\ n_i=1\end{cases}$.

□

上述给出的命题4.5.1的证明并未涉及相容矩阵范数.倘若使用相容矩阵范数,我们便得如下结果.

(1) $\lim\limits_{k\to\infty}A^k=0\Leftrightarrow\rho(A)<1$;

(2) $\sum\limits_{k=0}^{\infty}A^k$ 收敛 $\Leftrightarrow\sum\limits_{k=0}^{\infty}A^k$ 绝对收敛 $\Leftrightarrow\rho(A)<1$.

证明: (1)⇒　由 Schur 分解 $U^*AU=\begin{pmatrix}\lambda_1 & \cdots & *\\ & \ddots & \vdots\\ & & \lambda_n\end{pmatrix}$ 知 $\lim\limits_{k\to\infty}\lambda_i^k=0$, $i=1,2,\cdots,n$.因此, $|\lambda_i|<1$,即有 $\rho(A)<1$.

⇐　当 $\rho(A)<1$ 时, $\frac{1}{2}(1-\rho(A))>0$.于是存在 $\mathbb{C}^{n\times n}$ 上的相容矩阵范数 $\|\cdot\|$,使得

$$\|A\|\leqslant\rho(A)+\frac{1}{2}(1-\rho(A))=\frac{1}{2}(1+\rho(A))<1$$

结果 $\|A^k\|\leqslant\|A\|^k\to0$.再由定理4.4.1之(2)得 $A^k\to0$.

(2) $\sum\limits_{k=0}^{\infty}A_k$ 收敛 $\Rightarrow\lim\limits_{k\to\infty}A^k=0\Rightarrow\rho(A)<1$.当 $\rho(A)<1$ 时,取相容矩阵范数 $\|\cdot\|$ 使得 $\|A\|<1$,于是 $\sum\limits_{k=0}^{\infty}\|A^k\|\leqslant\sum\limits_{k=0}^{\infty}\|A\|^k$ 收敛.因此 $\sum\limits_{k=0}^{\infty}A^k$ 绝对收敛.

□

例 4.5.2　设 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 上的算子矩阵范数,则有

(1) $\|I_n\|=1$;

(2)若 $\|A\|<1$,则 I_n-A 可逆,且 $\|(I_n-A)^{-1}\|\leqslant\frac{1}{1-\|A\|}$, $\|(I_n-A)^{-1}-I\|\leqslant$

$\dfrac{\|A\|}{1-\|A\|}$;

(3)当 A 可逆时,$\|A^{-1}\| \geqslant \|A\|^{-1}$, $\|A^{-1}\|^{-1} = \inf\limits_{x\neq 0}\dfrac{\|Ax\|}{\|x\|}$.

证明:(1)显然.

(2)$\|A\|<1$,因此级数 $\sum\limits_{k=0}^{\infty}A^k$ 绝对收敛,其部分和 S_m 满足

$$(I_n - A)S_m = I_n - A^{m+1}$$
$$(I_n - A)\lim_{m\to\infty} S_m = I_n$$

这表明

$$(I_n - A)^{-1} = \lim_{m\to\infty} S_m$$

$$\begin{aligned}\|(I_n - A)^{-1}\| &= \|\lim_{m\to\infty} S_m\| = \lim_{m}\|S_m\| \\ &\leqslant \lim_{m\to\infty}(1+\|A\|+\cdots+\|A\|^m) \\ &= \lim_{m\to\infty}\frac{1-\|A\|^{m+1}}{1-\|A\|} \\ &= \frac{1}{1-\|A\|}\end{aligned}$$

注意到

$$(I_n - A)[(I_n - A)^{-1} - I_n] = A$$

因此

$$(I_n - A)^{-1} - I = (I_n - A)^{-1}A$$

结果

$$\|(I_n - A)^{-1} - I\| \leqslant \frac{\|A\|}{1-\|A\|}$$

(3)$1 = \|I_n\| = \|AA^{-1}\| \leqslant \|A\|\ \|A^{-1}\|$ 得

$$\|A^{-1}\| \geqslant \|A\|^{-1}$$

再由

$$\|A^{-1}\| = \sup_{x\neq 0}\frac{\|A^{-1}x\|}{\|x\|}$$

知

$$\|A^{-1}\|^{-1} = \inf_{x\neq 0}\frac{\|x\|}{\|A^{-1}x\|} = \inf_{x\neq 0,\, y=A^{-1}}\frac{\|Ay\|}{y}$$

□

作为矩阵序列的应用,下面证明非负矩阵论中著名的 Perron 定理.在矩阵代数中,称一个矩阵 $A=(a_{ij})$为**正矩阵**是指:A 的每个元素 a_{ij}都大于零,记作 $A>0$.称 A 为**非负矩阵**是指A 的每个元素 a_{ij}都大于等于零,记作 $A\geqslant 0$.特别需要指出的是,读者要注意此处的“$\geqslant$”、“$>$”和本书中命题 1.3.4 中出现的“$\leqslant$”之间的区别 . 首先给出以下引理.

引理 4.5.1 设 n 阶方阵$A>0$. p 为关于向量x 的不等式组

$$\begin{cases} x \geqslant 0 \\ (A - \rho(A)I)x \geqslant 0 \end{cases} \tag{4.5.2}$$

的一个非零解.则有

(1)$p > 0$,且 $Ap = \rho(A)p, \rho(A) > 0$;

(2)不等式(4.5.2)的解集为$\{kp \mid k \geqslant 0\}$.

证明:(1)当$(A - \rho(A)I)p = 0$时,直接知 $Ap = \rho(A)p$.当 $p \neq 0$ 又知,$Ap > 0$.因此有 $\rho(A) > 0, p > 0$.因此只需证$(A - \rho(A)I) \neq 0$. 现若$(A - \rho(A)I)p \neq 0$,于是 $A(A - \rho(A)I)p > 0$,这归功于 $A > 0$.结果存在 $\varepsilon > 0$,使得

$$A(A - \rho(A)I)p \geqslant \varepsilon Ap > 0$$

亦即有

$$A^2 p \geqslant \varepsilon Ap + A\rho(A)p = (\varepsilon + \rho(A))Ap > 0$$

令 $B = \dfrac{A}{\varepsilon + \rho(A)}$,于是 $\rho(B) < 1, \lim\limits_{k \to \infty} B^k = 0$.此时上式变为

$$BAp \geqslant Ap > 0$$

递推得

$$\begin{cases} B^2 Ap \geqslant Ap > 0 \\ \qquad \vdots \\ B^k Ap \geqslant Ap > 0 \\ \qquad \vdots \end{cases}$$

由 $\lim\limits_{k \to \infty} B^k = 0$,得 $0 \geqslant Ap \geqslant 0$,矛盾.(1)获证.

(2)当 q 也满足(4.5.2)式时,分两种情况考虑.情况一是 $q = 0$.此时结论显然.情况二是 $q \neq 0$,此时由(1)得 $q > 0, Aq = \rho(A)q$.

取 $k = \max\limits_{1 \leqslant i \leqslant n} \dfrac{q_i}{p_i}, p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}, q = \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}$,则

$$\begin{cases} (A - \rho(A)I)(kp - q) = 0 \\ kp - q \geqslant 0 \end{cases}$$

仍由(1)知,$kp - q > 0$,与 k 的选取矛盾.引理获证.

□

定理 4.5.2(Perron 定理)　设 $A \in \mathbb{R}^{n \times n}, A > 0$.于是有

(1)$\rho(A)$是 A 的正特征值,并且存在相应于它的正特征向量;

(2)A 的特征值中,模为 $\rho(A)$的特征值只有一个,它为 $\rho(A)$;

(3)$\rho(A)$的代数重数为 1;

(4)若 $B - A \geqslant 0$,则 $\rho(B) \geqslant \rho(A)$.进一步,当 $B \geqslant A$ 但 $B \neq A$ 时,有 $\rho(B) > \rho(A)$.

证明:(1)设 $Ax = \mu x$,其中 $0 \neq x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{C}^{n \times 1}, |\mu| = \rho(A)$.取绝对值后得

$$A|x| \geqslant \rho(A)|x|$$

$$(A-\rho(A)I)|x|\geqslant 0$$

其中,$0\neq|x|=\begin{pmatrix}|x_1|\\ \vdots\\ |x_n|\end{pmatrix}\geqslant 0$.由引理4.5.1之(1)知,$|x|>0,A|x|=\rho(A)|x|,\rho(A)>0$.

(2)设 $\lambda\in\operatorname{spec}A,|\lambda|=\rho(A)$.取相应于 λ 的特征向量 $0\neq x=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}\in\mathbb{C}^{n\times 1}$,于是有 $Ax=\lambda x\Rightarrow A|x|\geqslant\rho(A)|x|\Rightarrow A|x|=\rho(A)|x|=|Ax|$.于是再次由引理4.5.1之(1)得,$|x|>0$.现将 x_j 表示成 $x_j=|x_j|e^{i\theta_j},j=1,2,\cdots,n$.于是由 $A|x|=|Ax|$得

$$\left|\sum_j a_{ij}\cdot|x_j|\cdot e^{i\theta_j}\right|=\sum_j a_{ij}|x_j|$$

这表明:

$$\theta_1=\cdots=\theta_n=\theta\text{ 为常数}$$

此时再由 $A|x|=\rho(A)|x|$得$Ax=\rho(A)x$.注意到 $Ax=\lambda x$,于是 $\lambda=\rho(A)$如所需.

(3)首先 $\rho(A)$的几何重数为1.这是因为:对 $\rho(A)$的任何特征向量 x,$Ax=\rho(A)x$ 导致$A|x|\geqslant|Ax|=\rho(A)|x|$,因此又由引理4.5.1得 $A|x|=\rho(A)|x|=|Ax|$.重复(2)中的推导得 $x=|x|e^{i\theta}$,但$|x|$可统一用$\rho(A)$的一个正特征向量 p 来线性表示.这样,x 也可用该p 来线性表示.此即 $\rho(A)$的几何重数为1.也就是,A 关于$\rho(A)$的 Jordan 块只有1块.

倘若 A 关于$\rho(A)$的 Jordan 块不是1阶的,则可取可逆阵 $P=(p_1,p_2,\cdots)$使得

$$A(p_1,p_2,\cdots)=(p_1,p_2,\cdots)\begin{pmatrix}\rho(A)&1&&\\&\rho(A)&\ddots&\\&&\ddots&*\\&&&*\end{pmatrix}$$

取定 A 的相应于$\rho(A)$的一个正特征向量 q_1,由 $\rho(A)$的几何重数是1,于是可适当选取参数 k 使得$kp_1=q_1$.记 $q_2=\operatorname{Re}(kp_2)$,于是

$$A(kp_1,kp_2,\cdots)=(kp_1,kp_2,\cdots)\begin{pmatrix}\rho(A)&1&&\\&\rho(A)&\ddots&\\&&\ddots&*\\&&&*\end{pmatrix}$$

从而得

$$\begin{cases}Aq_1=\rho(A)q_1\\ Aq_2=\rho(A)q_2+q_1\\ q_1,q_2\text{ 线性无关},q_1>0\end{cases}$$

现在让 τ 为足够大的正数使得$\tau q_1+q_2>0$,则有

$$A(\tau q_1+q_2)=\rho(A)(\tau q_1+q_2)+q_1$$
$$A(\tau q_1+q_2)\geqslant\rho(A)(\tau q_1+q_2)$$

再一次利用引理 4.5.1,有 τq_1+q_2 也为相应于 $\rho(A)$ 的特征向量.仍由 $\rho(A)$ 的几何重数是 1 知:τq_1+q_2 可表示为 $\tau q_1+q_2=tq_1$.结果 q_1,q_2 线性相关.矛盾.

(4) $B-A\geqslant 0$ 表明 $B\geqslant A$.于是 $\forall\varepsilon>0,\rho\left(\frac{B}{\rho(B)+\varepsilon}\right)<1$,从而

$$\lim_{k\to\infty}\left(\frac{B}{\rho(B)+\varepsilon}\right)^k=0$$

注意到

$$\left(\frac{B}{\rho(B)+\varepsilon}\right)^k\geqslant\left(\frac{A}{\rho(B)+\varepsilon}\right)^k\geqslant 0$$

于是

$$\lim_{k\to\infty}\left(\frac{A}{\rho(B)+\varepsilon}\right)^k=0$$

亦即

$$\frac{\rho(A)}{\rho(B)+\varepsilon}<1$$

故有

$$\rho(A)<\rho(B)+\varepsilon$$

让 $\varepsilon\to 0$ 得

$$\rho(A)\leqslant\rho(B)$$

进一步,倘若 $B\geqslant A$ 但 $B\neq A$,我们断定 $\rho(A)\neq\rho(B)$,否则,取 $x>0$ 满足 $Ax=\rho(A)x$.于是 $Ax=\rho(B)x,(B-\rho(B)I)x=(B-A)x\geqslant 0$.又由引理 4.5.1 知:$Bx=\rho(B)x$.这表明 $(B-A)x=(B-\rho(B)I)x=0$.明显,由 $x>0$ 及 $B\neq A$ 知,$(B-A)x\neq 0$.因此断言真.

□

在矩阵论中,人们也往往通过矩阵范数来刻画矩阵的最佳逼近问题.例如,下例解决了在 Frobenius 范数意义下,所有秩为 r 的 $m\times n$ 阶矩阵中,如何寻找给定矩阵 A 的一个最佳逼近问题.

例 4.5.3　设 $A\in\mathbb{C}_l^{m\times n},m\geqslant n,l\geqslant r,A$ 的奇异值为 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant 0$.记 $S=\begin{pmatrix}\sigma_1&&\\&\ddots&\\&&\sigma_n\end{pmatrix}_{n\times n},\Sigma=\begin{pmatrix}\sigma_1&&\\&\ddots&\\&&\sigma_r\end{pmatrix}_{r\times r}$,$A$ 的奇异值分解为 $A=U\begin{pmatrix}S\\0\end{pmatrix}V^*$.又记 $A_0=U\begin{pmatrix}\Sigma&0\\0&0\end{pmatrix}_{m\times n}V^*$,则有 $\|A-A_0\|_F=\min\limits_{\text{秩}B=r}\|A-B\|_F$.

证明: 首先,断言 $\min\limits_{\text{秩}B\leqslant r}\|A-B\|_F$ 存在.事实上,取 $\{B_i:\text{秩 }B_i\leqslant r\}$ 使得 $\|A-B_i\|_F\to\inf\limits_{\text{秩}B\leqslant r}\|A-B\|_F$,于是 $\exists$ 子列 $\{B_{n_i}:\text{秩 }B_{n_i}\leqslant r\}$,使得 $\lim\limits_{i\to\infty}B_{n_i}$ 存在并记为 B_0,此时秩 $B_0\leqslant r$.结果 $\|A-B_0\|_F=\lim\limits_{i\to\infty}\|A-B_{n_i}\|_F=\inf\limits_{\text{秩}B\leqslant r}\|A-B\|_F$.

其次,设 B_0 的奇异值分解为

$$B_0 = U_1 \begin{pmatrix} C_{11} & 0 \\ 0 & 0 \end{pmatrix} V_1^*$$

其中 U_1, V_1 为酉阵，$C_{11} = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_s \end{pmatrix}_{s \times s}$，$\gamma_1 \geqslant \cdots \geqslant \gamma_s$ 为 B_0 的非零奇异值.我们来看 $\| A - B_0 \|_F$ 能成为 $\inf\limits_{\text{秩} B \leqslant r} \| A - B \|_F$ 所需的条件.为使 $\| A - B_0 \|_F = \inf\limits_{\text{秩} B \leqslant r} \{ \| A - B \|_F \}$，构造矩阵

$$C = U_1^* B_0 V_1 = \begin{pmatrix} C_{11} & 0 \\ 0 & 0 \end{pmatrix}_{m \times n}$$

$$D = U_1^* A V_1 = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$$

容易知道必须要有 $D_{11} = C_{11}, D_{12} = 0, D_{21} = 0$.否则以 $D_{12} \neq 0$ 为例，构造矩阵

$$M = \begin{pmatrix} C_{11} & D_{12} \\ 0 & 0 \end{pmatrix} \in \mathbb{F}_s^{m \times n}$$

则有 $\| D - M \|_F < \| D - C \|_F$.进而

$$\| D - M \|_F = \| U_1 (D - M) V_1^* \|_F = \| A - U_1 M V_1^* \|_F < \| D - C \|_F = \| A - B_0 \|_F$$

这与 $\| A - B_0 \|_F$ 达到最小值矛盾.类似地可证 $D_{21} = 0, D_{11} = C_{11}$.据此 D 实际上变为：

$$D = \begin{pmatrix} C_{11} & 0 \\ 0 & D_{22} \end{pmatrix}$$

注意到 D 的奇异值与 A 的奇异值相同，它为 C_{11} 的奇异值与 D_{22} 的奇异值的并集.又要使 $\| D - C \|_F = \| A - B_0 \|_F$ 达到最小，C_{11} 的奇异值只能为 A 的前 s 个奇异值，因此有

$$\inf_{\text{秩} B \leqslant r} \| A - B \|_F = \left(\sum_{i = s+1}^{n} \sigma_i^2 \right)^{\frac{1}{2}}$$

最后直接验证可得

$$\| A - A_0 \|_F = \left(\sum_{i = r+1}^{n} \sigma_i^2 \right)^{\frac{1}{2}}$$

注意到 $s \leqslant r$，因此 B_0 可取为 A_0，从而 $\| A - A_0 \|_F = \min\limits_{\text{秩} B = r} \| A - B \|_F$.

□

4.6 计算 A^+ 的几种方法

第 3 章中介绍了利用满秩分解计算加号逆的方法.利用矩阵序列和矩阵级数，本节进一步介绍计算 A^+ 的几种方法.

首先，利用 A 的奇异值分解，我们可得关于 A^+ 的 Lagrange - Sylvester 公式及 Neumann 展式.

定理 4.6.1(Lagrange – Sylvester 公式)　设 $A\in\mathbb{C}_s^{m\times n}$, $s\geqslant 1$, $\sigma_1,\cdots,\sigma_r$ 为 A 的所有互异非零奇异值.则

$$A^+ = \sum_{i=1}^{r}\sigma_i^{-2}\frac{\prod\limits_{j=1(j\neq i)}^{r}(A^*A-\sigma_j^2 I)}{\prod\limits_{j=1(j\neq i)}^{r}(\sigma_i^2-\sigma_j^2)}A^* \tag{4.6.1}$$

证明: 设 A^*A 的谱分解式为 $A^*A=\sum\limits_{i=1}^{r}\sigma_i^2 p_i$, 于是 $(A^*A)^+=\sum\limits_{i=1}^{r}\sigma_i^{-2}p_i$, 并令

$$g_i(\lambda) = \prod_{j=1(j\neq i)}^{r}(\lambda-\sigma_j^2)$$

于是

$$g_i(A^*A) = \sum_{j=1}^{r}g_i(\sigma_j^2)p_j = g_i(\sigma_i^2)p_i$$

从而

$$p_i = \frac{g_i(A^*A)}{g_i(\sigma_i^2)}$$

注意到(见习题三)

$$A^+ = (A^*A)^+A^* = \sum_{i=1}^{r}\sigma_i^{-2}\frac{\prod\limits_{\substack{j\neq i\\ j=1}}^{r}(A^*A-\sigma_j^2 I)}{\prod\limits_{j=1(j\neq i)}^{r}(\sigma_i^2-\sigma_j^2)}A^*$$

□

定理 4.6.2(Neumann 展式)　设 $A\in\mathbb{C}_r^{m\times n}$, $r\geqslant 1$, $\sigma_1,\cdots,\sigma_r$ 为 A 的非零奇异值.令 $c=\max\limits_{1\leqslant i\leqslant r}\sigma_i^2$, $a\in\left(0,\dfrac{2}{c}\right)$, 则有

$$A^+ = a\sum_{k=0}^{\infty}(I-aA^*A)^kA^* \tag{4.6.2}$$

证明: 取 A 的奇异值分解 $A=UDV^*$, 其中

$$D=\begin{pmatrix}S & 0\\ 0 & 0\end{pmatrix}_{m\times n},\quad S=\begin{pmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_r\end{pmatrix}_{r\times r}$$

于是 $A^*A=V\begin{pmatrix}S^2 & 0\\ 0 & 0\end{pmatrix}V^*$.故有

$$(I-aA^*A)^k = V\left[I-a\begin{pmatrix}S^2 & 0\\ 0 & 0\end{pmatrix}\right]^kV^*$$

$$\begin{aligned}a(I-aA^*A)^kA^* &= Va\left[I-a\begin{pmatrix}S^2 & 0\\ 0 & 0\end{pmatrix}\right]^kV^*VD^*U^*\\ &= Va\left[I-\begin{pmatrix}aS^2 & 0\\ 0 & 0\end{pmatrix}\right]^kD^*U^*\end{aligned}$$

观察矩阵 $W \triangleq a\sum_{k=0}^{\infty}\left[I_n - \begin{pmatrix} aS^2 & 0 \\ 0 & 0 \end{pmatrix}\right]^k D^*$，$W$ 为对角阵,其非零对角元为

$$a\sum_{k=0}^{\infty}(1 - a\sigma_i^2)^k\sigma_i = a\sigma_i\sum_{k=0}^{\infty}(1 - a\sigma_i^2)^k = \frac{1}{\sigma_i}$$

因此

$$W = D^+$$

$$A^+ = VD^+ U^* = a\sum_{k=0}^{\infty}(I - aA^* A)^k A^*$$

□

注意到 $(I - aA^* A)^k A^* = A^*(I - aAA^*)^k$，因此(4.6.2)式又可写为

$$A^+ = a\sum_{k=0}^{\infty}A^*(I - aAA^*)^k$$

定理 4.6.3 在定理 4.6.2 条件下,任取正整数 $p\geqslant 2$,记

$$A_p^+ = a\left\{\prod_{k=0}^{p-1}[I + (I - aA^* A)^{2^k}]\right\}A^*$$

则有

$$\| A^+ - A_p^+ \|_2 = \max_i \frac{(1 - a\sigma_i^2)^{2^p}}{\sigma_i}$$

证明: 由归纳法易证下列欧拉恒等式

$$(1 + x)\prod_{k=1}^{p-1}(1 + x^{2^k}) = \sum_{k=0}^{2^p-1}x^k$$

于是得

$$[I + (I - aA^* A)]\prod_{k=1}^{p-1}[I + (I - aA^* A)^{2^k}] = \sum_{k=0}^{2^p-1}(I - aA^* A)^k$$

因此

$$A^+ - A_p^+ = a\sum_{k=2^p}^{\infty}(I - aA^* A)^k A^* = aV\sum_{k=2^p}^{\infty}\left(I_n - \begin{pmatrix} aS^2 & 0 \\ 0 & 0 \end{pmatrix}\right)^k D^* U^*$$

再由谱范数为酉不变范数,得

$$\| A^+ - A_p^+ \|_2 = \left\| a\sum_{k=2^p}^{\infty}\left(I_n - \begin{pmatrix} aS^2 & 0 \\ 0 & 0 \end{pmatrix}\right)^k D^* \right\|_2 = \max_i \frac{(1 - a\sigma_i^2)^{2^p}}{\sigma_i}$$

□

下面给出计算 A^+ 的两种迭代方法.迭代方法的目的是获取一个收敛于 A^+ 的矩阵序列 $\{X_k\}$,迭代过程的收敛速度取决于相应的残差序列:

$$R_k = AA^+ - AX_k, k = 0,1,2,\cdots$$

定理 4.6.4 设 $0\neq A\in \mathbb{C}^{m\times n}$.取初值 X_0 和残差初值 R_0 为

$$\begin{cases} X_0 = A^* B_0 A^*, \text{对某个 } B_0\in \mathbb{C}^{m\times n} \\ R_0 = AA^+ - AX_0 \\ \rho(R_0) < 1 \end{cases}$$

则矩阵序列

$$\begin{cases} X_0 = A^* B_0 A^* \\ T_k = I - AX_k \\ X_{k+1} = X_k + X_0 T_k \\ k = 0, 1, \cdots \end{cases}$$

当 $k \to \infty$ 时，$X_k \to A^+$，相应的残差 $R_k = AA^+ - AX_k$ 满足

$$\| R_{k+1} \| \leqslant \| R_0 \| \, \| R_k \|, k = 0, 1, \cdots$$

其中 $\| \cdot \|$ 为 $\mathbb{C}^{m \times m}$ 上的任意相容矩阵范数.

证明: 直接验证知 $X_0 AA^+ = A^* B_0 A^* AA^+ = A^* B_0 (AA^+ A)^* = X_0$. 因此有

$$X_{k+1} = X_k + X_0(I - AX_k) = X_k + X_0 R_k$$

$$\begin{aligned} R_{k+1} &= AA^+ - AX_{k+1} = AA^+ - AX_k - AX_0 R_k \\ &= R_k - AX_0 R_k = AA^+ R_k - AX_0 R_k = R_0 R_k \end{aligned}$$

$$\| R_{k+1} \| \leqslant \| R_0 \| \, \| R_k \|$$

$$R_{k+1} = R_0 R_k = R_0^2 R_{k-1} = \cdots = R_0^{k+2}$$

$$k = 0, 1, 2, \cdots$$

由 $\rho(R_0) < 1$ 知

$$\lim_{k \to \infty} R_k = 0, \text{ i.e. } \lim_{k \to \infty} AX_k = AA^+$$

接下来，证明 $\lim\limits_{k \to \infty} X_k$ 存在，且 $\lim\limits_{k \to \infty} X_k = A^+$.

注意到 $\rho(R_0) < 1$，故可取相容矩阵范数 $\| \cdot \|'$，使得 $\rho(R_0) \leqslant \| R_0 \|' < 1$. 再由矩阵范数的等价性又可取正数 α, β，使得 $\forall M \in \mathbb{C}^{n \times n}$ 均有

$$\alpha \| M \|' \leqslant \| M \| \leqslant \beta \| M \|'$$

于是

$$\| X_{k+1} - X_k \| = \| X_0 R_k \| = \| X_0 R_0^{k+1} \| \leqslant \beta \| X_0 R_0^{k+1} \|' \leqslant \beta \| X_0 \|' (\| R_0 \|')^{k+1}$$

结果 $\sum\limits_{k=0}^{\infty} (X_{k+1} - X_k)$ 收敛，$\lim\limits_{k \to \infty} X_k$ 存在. 若记 $X = \lim\limits_{k \to \infty} X_k$，于是 $AX = AA^+$，$X \in A\{1,3\}$. 为证 $X \in A\{2,4\}$，我们证明 $X^* XA = X^*$ 成立即可.

考虑递推式:

$$X_k = X_{k-1} + X_0 R_0^k = X_{k-2} + X_0 R_0^{k-1} + X_0 R_0^k = \cdots = X_0(I + R_0 + \cdots + R_0^k)$$

所以

$$X = X_0(I - R_0)^{-1}$$

$$X_0^* XA = AB_0^* AXA = AB_0^* AA^+ A = AB_0^* A = X_0^*$$

$$X^* XA = [(I - R_0)^*]^{-1} X_0^* XA = [(I - R_0)^*]^{-1} X_0^* = X^*$$

最后，$X = A^+$. 证毕.

□

定理 4.6.5　设 A，初值 X_0 及残差 R_0 满足定理 4.6.4 的条件，p 为大于等于 2 的任意整数，则迭代格式

$$\begin{cases} X_0 = A^* B_0 A^* \\ T_k = I - AX_k \\ X_{k+1} = X_k(I + T_k + T_k^2 + \cdots + T_k^{p-1}) \\ k = 0,1,\cdots \end{cases}$$

中，$\lim_{k\to\infty} X_k = A^+$，并且相应的残差序列$\{R_k = AA^+ - AX_k\}$满足

$$\| R_{k+1} \| \leqslant \| R_k \|^p, k = 0,1,\cdots$$

其中$\|\cdot\|$为$\mathbb{C}^{m\times m}$上的任意相容矩阵范数.

证明：由 $X_0 = A^* B_0 A^*$ 出发，容易看出，每个 X_k 均可表示为$X_k = A^* B_k A^*$ 的形式，其中 B_k 为 $\mathbb{C}^{m\times n}$中的某元.因此

$$X_k AA^+ = X_k$$

$$X_k R_k = X_k(AA^+ - AX_k) = X_k(I - AX_k) = X_k T_k$$

$$\begin{aligned} X_k T_k^2 &= X_k(I - AX_k)(I - AX_k) \\ &= (I - X_k A) X_k (I - AX_k) \\ &= (I - X_k A) X_k R_k \\ &= X_k(I - AX_k) R_k \\ &= X_k R_k^2 \\ &\;\;\vdots \\ X_k T_k^{p-1} &= X_k(I - AX_k) T_k^{p-2} \\ &= (I - X_k A) X_k T_k^{p-2} \\ &= (I - X_k A) X_k R_k^{p-2} \\ &= X_k(I - AX_k) R_k^{p-2} \\ &= X_k R_k^{p-1} \end{aligned}$$

结果

$$X_{k+1} = X_k(I + T_k + T_k^2 + \cdots + T_k^{p-1}) = X_k(I + R_k + R_k^2 + \cdots + R_k^{p-1})$$

从而

$$\begin{aligned} R_{k+1} &= AA^+ - AX_{k+1} \\ &= AA^+ - AX_k(I + R_k + R_k^2 + \cdots + R_k^{p-1}) \\ &= AA^+ - AX_k - AX_k(R_k + R_k^2 + \cdots + R_k^{p-1}) \\ &= R_k - AX_k(R_k + R_k^2 + \cdots + R_k^{p-1}) \end{aligned}$$

注意到 $AA^+ R_k = R_k$，于是对所有的 $q = 1,2,\cdots,p-1$，均有

$$R_k^q - AX_k R_k^q = AA^+ R_k^q - AX_k R_k^q = R_k^{q+1}$$

故

$$\begin{aligned} R_{k+1} &= R_k - AX_k(R_k + R_k^2 + \cdots + R_k^{p-1}) \\ &= R_k - AX_k R_k - AX_k(R_k^2 + \cdots + R_k^{p-1}) \\ &= R_k^p \end{aligned}$$

此时

$$\| R_{k+1} \| = \| R_k^p \| \leqslant \| R_k \|^p$$

$$\lim_{k\to\infty} R_k = 0$$

$$\lim_{k\to\infty} AX_k = AA^+$$

最后类似于定理 4.6.4 的证明，有 $\lim_{k\to\infty} X_k$ 存在，且 $\lim_{k\to\infty} X_k = A^+$.

□

表面上看定理 4.6.4 的迭代算法的收敛速度是一阶的，而定理 4.6.5 给的迭代算法则是 $p(\geqslant 2)$ 阶的，似乎定理 4.6.5 给的迭代算法要好些. 但仔细观察不难发现，定理 4.6.5 中的每次迭代所需要的计算却比定理 4.6.4 中的要多. 因此，在实际应用中须考虑选择适当的阶次 p，使得计算量较小.

关于 A^+ 的计算，Greville 另辟蹊径地给出了一种有限次的递推算法，现称为 **Greville 方法**. 其主要思路是：首先将 A_k 分块为

$$A_k = [A_{k-1} \vdots \alpha_k], \quad k = 2,3,\cdots,n$$

其中 α_k 为 A 的第 k 列，A_{k-1} 为 A 的前 $k-1$ 列. 然后通过 A_{k-1}^+ 给出 A^+ 的准确表示. 为方便叙述，令

$$d_k = A_{k-1}^+ \alpha_k$$

$$c_k = \alpha_k - A_{k-1} d_k = (I - A_{k-1} A_{k-1}^+) \alpha_k$$

于是 Greville 方法便可被概括为如下定理：

定理 4.6.6　设 $A \in \mathbb{C}^{m\times n}$，则 $A_k (2 \leqslant k \leqslant n)$ 的加号逆为

$$A_k^+ = [A_{k-1} \vdots \alpha_k]^+ = \begin{bmatrix} A_{k-1}^+ - d_k b_k^* \\ b_k^* \end{bmatrix}$$

其中

$$b_k^* = \begin{cases} c_k^+, & c_k \neq 0 \text{ 时} \\ \dfrac{d_k^* A_{k-1}^+}{1 + d_k^* d_k}, & c_k = 0 \text{ 时} \end{cases}$$

证明： 将 $A_k^+ = [A_{k-1} \vdots \alpha_k]^+$ 分块为

$$A_k^+ = \begin{bmatrix} B_k \\ b_k^* \end{bmatrix}$$

其中 b_k^* 为 A_k^+ 的第 k 行. 于是

$$A_k A_k^+ = A_{k-1} B_k + \alpha_k b_k^* \tag{4.6.3}$$

由于 $A_k A_k^+$ 与 $A_{k-1}^+ A_{k-1}$ 均是正交投影，因此容易验证下列两等式成立：

$$A_{k-1}^+ A_k A_k^+ = A_{k-1}^+ \tag{4.6.4}$$

$$A_{k-1}^+ A_{k-1} B_k = B_k \tag{4.6.5}$$

其中式(4.6.4)归功于

$$N(A_{k-1}^+) = N(A_{k-1}^*) \supseteq N(A_k^*) = N(A_k A_k^+)$$

式(4.6.5)归功于

$$\begin{cases}\begin{bmatrix}B_k\\b_k^*\end{bmatrix}=A_k^+[A_{k-1}\ \vdots\ \alpha_k]A_k^+=\begin{bmatrix}A_{k-1}^*\\\alpha_k^*\end{bmatrix}(A_k^+)^*A_k^+\\R(B_k)\subseteq R(A_{k-1}^*)=R(A_{k-1}^+)=R(A_{k-1}^+A_{k-1})\end{cases}$$

现对式(4.6.3)两端左乘 A_{k-1}^+得

$$A_{k-1}^+=A_{k-1}^+A_{k-1}B_k+A_{k-1}^+\alpha_kb_k^*=B_k+A_{k-1}^+\alpha_kb_k^*$$

$$B_k=A_{k-1}^+-A_{k-1}^+\alpha_kb_k^*=A_{k-1}^+-d_kb_k^*$$

$$[A_{k-1}\ \vdots\ \alpha_k]^+=\begin{bmatrix}A_{k-1}^+-d_kb_k^*\\b_k^*\end{bmatrix}$$

遗留下来的工作是确定 b_k^*.就 c_k 分两种情形讨论.

第一种情形 $c_k\neq0$.此时由 $B_k=A_{k-1}^+-d_kb_k^*$ 得

$$\begin{aligned}A_kA_k^+&=A_{k-1}B_k+\alpha_kb_k^*\\&=A_{k-1}A_{k-1}^++(\alpha_k-A_{k-1}d_k)b_k^*\\&=A_{k-1}A_{k-1}^++c_kb_k^*\end{aligned}$$

移项知:$c_kb_k^*$ 是自伴阵,秩不超过1,因此它可分解为

$$c_kb_k^*=\xi\beta\beta^*,\beta\in\mathbb{C}^{m\times1},\xi\in\mathbb{R}$$

故

$$c_kb_k^*\beta=\xi\beta\|\beta\|^2$$

当 $b_k^*=0$ 时,b_k^* 可表为 $b_k^*=0\cdot c_k^*$.

当 $b_k^*\neq0$ 时,由 $c_k\neq0$ 知 $\xi\neq0,\beta\neq0$,从而 β 可写成 $\beta=sc_k$,进而 $b_k^*=\dfrac{\xi c_k^*\beta\bar{s}c_k^*}{\|c_k\|^2}=\delta c_k^*,\delta=\dfrac{\xi|s|^2\|c_k\|^2}{\|c_k\|^2}=\xi|s|^2$ 为一实数.总之有

$$b_k^*=\delta c_k^*,\delta\in\mathbb{R}$$

现在根据 A_k 的分块表示及 c_k 的定义得

$$\begin{aligned}A_k&=A_kA_k^+A_k=(A_{k-1}A_{k-1}^++c_kb_k^*)[A_{k-1}\ \vdots\ \alpha_k]\\&=[A_{k-1}+c_kb_k^*A_{k-1}\ \vdots\ A_{k-1}A_{k-1}^+\alpha_k+c_kb_k^*\alpha_k]\\&=[A_{k-1}+c_kb_k^*A_{k-1}\ \vdots\ \alpha_k-c_k+(b_k^*\alpha_k)c_k]\end{aligned}$$

对比 A_k 的分块表示$A_k=[A_{k-1}\ \vdots\ \alpha_k]$,于是有 $c_k=(b_k^*\alpha_k)c_k,b_k^*\alpha_k=1$.进而

$$\begin{aligned}1=b_k^*\alpha_k&=\delta c_k^*\alpha_k\\&=\delta\alpha_k^*(I-A_{k-1}A_{k-1}^+)\alpha_k\\&=\delta\alpha_k^*(I-A_{k-1}A_{k-1}^+)(I-A_{k-1}A_{k-1}^+)\alpha_k\\&=\delta c_k^*c_k\end{aligned}$$

结果

$$\delta=\frac{1}{\|c_k\|^2},\ b_k^*=\delta c_k^*=c_k^+$$

第二种情形　$c_k = 0$. 此时 $\alpha_k = A_{k-1}A_{k-1}^{+}\alpha_k$, $\alpha_k \in R(A_{k-1})$, 因而

$$R(A_k) = R(A_{k-1})$$

结果有

$$N(b_k^*) \supseteq N(A_k^+) = N(A_k^*) = N(A_{k-1}^*) = N(A_{k-1}^+) = N(A_{k-1}A_{k-1}^+)$$

进而得

$$b_k^*(I - A_{k-1}A_{k-1}^+) = 0$$

$$b_k^* = b_k^* A_{k-1}A_{k-1}^+.$$

再由 $A_k = [A_{k-1} \vdots \alpha_k]$, $[A_{k-1} \vdots \alpha_k]^+ = \begin{bmatrix} A_{k-1}^+ - d_k b_k^* \\ b_k^* \end{bmatrix}$, 有

$$A_k^+ A_k = \begin{bmatrix} A_{k-1}^+ A_{k-1} - d_k b_k^* A_{k-1} & A_{k-1}^+ \alpha_k - d_k b_k^* \alpha_k \\ b_k^* A_{k-1} & b_k^* \alpha_k \end{bmatrix}$$

$$= \begin{bmatrix} A_{k-1}^+ A_{k-1} - d_k b_k^* A_{k-1} & (1 - b_k^* \alpha_k) d_k \\ b_k^* A_{k-1} & b_k^* \alpha_k \end{bmatrix}$$

让 $a = b_k^* \alpha_k$ 为某一复数,注意到 a 为 $A_k^+ A_k$ 的对角元,而 $A_k^+ A_k$ 为自共轭幂等阵,故又有

(1) a 为实数;

(2) $b_k^* A_{k-1} = (1-a) d_k^*$;

(3) $b_k^* = b_k^* A_{k-1} A_{k-1}^+ = (1-a) d_k^* A_{k-1}^+$;

(4) $a = b_k^* \alpha_k = (1-a) d_k^* A_{k-1}^+ \alpha_k = (1-a) d_k^* d_k$.

将(4)式两端同加上$(1-a)$得:

$$(1-a) + a = (1-a) + (1-a) d_k^* d_k = (1-a)(1 + d_k^* d_k)$$

$$(1-a) = \frac{1}{1 + d_k^* d_k}$$

这表明:

$$b_k^* = (1-a) d_k^* A_{k-1}^+ = \frac{1}{1 + d_k^* d_k} d_k^* A_{k-1}^+$$

整个定理证毕.

□

用 Greville 方法计算 Moore－Penrose 逆,在每次递推时加进一列.这种方法在实际应用中是方便的.倘若新矩阵是由旧矩阵添加一行而得,则新旧矩阵的加号逆之间完全可用上述对偶的形式建立它们之间的递推关系式.

习题四

1. 证明:矩阵的 Frobenius 范数是酉不变范数.

2. 若 $A \in \mathbb{R}^{n \times n}$ 是一个正矩阵,证明矩阵 $B = \frac{1}{\rho(A)} A$ 是幂收敛的.

3.设 $A\in\mathbb{C}^{n\times n}$，$\|\cdot\|$ 为任一种相容的矩阵范数.证明 $\rho(A)=\lim\limits_{n}\sqrt[n]{\|A^n\|}$.

4.让 $A\in\mathbb{C}^{m\times n}$，$B\in\mathbb{C}^{n\times q}$，$A$，$B$ 的奇异值分别为

$$\sigma_1(A)\geqslant\cdots\geqslant\sigma_n(A),\ \sigma_1(B)\geqslant\cdots\geqslant\sigma_n(B)$$

证明：(1) $\|A\|_F^2=\sum\limits_{i=1}^{n}\sigma_i^2(A)$；

(2)$\sigma_1(AB)\leqslant\sigma_1(A)\sigma_1(B)$，$\sigma_1(A+B)\leqslant\sigma_1(A)+\sigma_1(B)$，其中 $\sigma_1(AB)$，$\sigma_1(A+B)$分别表示 AB，$A+B$ 的最大奇异值.

5.设 $A\in\mathbb{C}^{n\times n}$，$Q\in\mathbb{C}^{n\times q}$且 $Q^*Q=I_q$，称方阵 $R_A(Q)=Q^*AQ$ 为矩阵A 关于 Q 的 Rayleigh 商.证明：$\forall\, x\in\mathbb{C}^{q\times 1}$，均有

$$\min_{B}\|AQx-QBx\|_2=\|AQx-QR_A(Q)x\|_2$$

6.求解双侧旋转问题：设 A，$B\in\mathbb{C}^{m\times n}$，求 $\max\limits_{U,V\text{酉阵}}\|A-UBV\|_F$.

7.证明 $\|AB\|_F\leqslant\|A\|_F\|B\|_F$，其中 A，$B\in\mathbb{C}^{n\times n}$.

第5章　矩阵函数及其计算

上一章中,我们用矩阵幂级数,对在开圆域内解析的复函数给出了它的矩阵函数的定义.本章解决的主要问题是:对于一般的复函数 $f(z)$,怎样自然而又合理地定义 $f(A)$(这里 A 是 n 阶复方阵)?

5.1　矩阵函数的定义

定义 5.1.1　设矩阵 $A\in\mathbb{C}^{n\times n}$ 的最小多项式为

$$m_A(\lambda)=(\lambda-\lambda_1)^{l_1}\cdots(\lambda-\lambda_t)^{l_t}$$

其中 $\lambda_1,\cdots,\lambda_t$ 各不相同.称集合 $\{(\lambda_i,l_i)\mid i=1,2,\cdots,t\}$ 为 A 的**简谱**,记为 Λ_A.

定义 5.1.2　设 $A\in\mathbb{C}^{n\times n}$,称函数 $f(\lambda)$ 在 A 的简谱 $\boldsymbol{\Lambda_A}$ **上确定**,是指

$$\begin{cases}f(\lambda_1),f'(\lambda_1),\cdots,f^{(l_1-1)}(\lambda_1)\\ \quad\vdots\\ f(\lambda_t),f'(\lambda_t),\cdots,f^{(l_t-1)}(\lambda_t)\end{cases}\tag{5.1.1}$$

均存在.若函数 $f(\lambda)$ 在 Λ_A 上确定,则记上述数组(5.1.1)为 $f(\Lambda_A)$.

由定义 5.1.1 及定义 5.1.2,易知如下事实.

定理 5.1.1　设 $A\in\mathbb{C}^{n\times n}$,$f(\lambda)$,$g(\lambda)$ 为复系数多项式,则有

$$f(A)=g(A)\Leftrightarrow f(\Lambda_A)=g(\Lambda_A)$$

证明: $f(A)=g(A)\Leftrightarrow m_A(\lambda)\mid(f(\lambda)-g(\lambda))$

$\Leftrightarrow(\lambda-\lambda_i)^{l_i}$ 至少是 $f(\lambda)-g(\lambda)$ 的 l_i 重因式 $(i=1,2,\cdots,t)$

$\Leftrightarrow f^{(j)}(\lambda_i)=g^{(j)}(\lambda_i)(j=0,1,\cdots,l_i-1,i=1,2,\cdots,t)$

$\Leftrightarrow f(\Lambda_A)=g(\Lambda_A)$

□

由此定理出发,我们给出矩阵函数的定义如下:

定义 5.1.3　设 $A\in\mathbb{C}^{n\times n}$,$f(\lambda)$ 为一个在 Λ_A 上确定的复函数.若存在复系数多项式 $p(\lambda)$,使得 $p(\Lambda_A)=f(\Lambda_A)$,则定义 $f(A)$ 为 $p(A)$.此时称 $p(\lambda)$ 为 $f(A)$ 的**定义多项式**.

显然,由定理5.1.1知,尽管 $f(A)$ 的定义多项式可能有多个,但 $f(A)$ 与可能有多种选择的多项式 $p(\lambda)$ 无关.当然,该定义中也还有一个问题需解决:即所要求的多项式 $p(\lambda)$ 是否存在?答案是肯定的.不仅如此,这种满足要求的 $f(A)$ 的定义多项式在次数小于

$\deg m_A(\lambda)$的约束下是唯一的.

定理 5.1.2 设 $A\in\mathbb{C}^{n\times n}$,$f(\lambda)$为一个在 Λ_A 上确定的复函数.则

(1)存在唯一的多项式 $p(\lambda)$,使 $\deg p(\lambda)<\deg m_A(\lambda)$,和 $p(\Lambda_A)=f(\Lambda_A)$;

(2)记 $f_i(\lambda)=\dfrac{f(\lambda)(\lambda-\lambda_i)^{l_i}}{m_A(\lambda)}$,则(1)中的 $p(\lambda)$为

$$p(\lambda)=m_A(\lambda)\sum_{i=1}^{t}\left[\frac{f_i(\lambda_i)}{(\lambda-\lambda_i)^{l_i}}+\cdots+\frac{f_i^{(l_i-1)}(\lambda_i)/(l_i-1)!}{\lambda-\lambda_i}\right]$$

证明:(1)记 $m=\deg m_A(\lambda)$.让 $p(\lambda)=a_1\lambda^{m-1}+\cdots+a_{m-1}\lambda+a_m$,其中 $a_1,\cdots,a_m$ 待定.由条件 $p(\Lambda_A)=f(\Lambda_A)$得 m 个线性方程.于是该方程组有唯一解等价于齐次线性方程 $p(\Lambda_A)=0$ 有唯一解.注意到 $p(\Lambda_A)=0$ 时有

$$\begin{cases}(\lambda-\lambda_1)^{l_1}\mid p(\lambda)\\ \qquad\vdots\\ (\lambda-\lambda_t)^{l_t}\mid p(\lambda)\end{cases}$$

从而 $m_A(\lambda)=(\lambda-\lambda_1)^{l_1}\cdots(\lambda-\lambda_t)^{l_t}\mid p(\lambda)$.由 $\deg p(\lambda)<m$ 知$p(\lambda)=0$.(1)获证.

(2)记

$$p_i(\lambda)=\frac{p(\lambda)(\lambda-\lambda_i)^{l_i}}{m_A(\lambda)}$$

$$f_i(\lambda)=\frac{f(\lambda)(\lambda-\lambda_i)^{l_i}}{m_A(\lambda)}$$

于是由 $p(\Lambda_A)=f(\Lambda_A)$得

$$p_i^{(j)}(\lambda_i)=f_i^{(j)}(\lambda_i),j=0,1,\cdots,l_i-1$$

注意到有理函数 $p_i(\lambda)$在 $\lambda=\lambda_i$ 的邻域内解析,故有函数

$$\frac{p_i(\lambda)-\displaystyle\sum_{j=0}^{l_i-1}\frac{p_i^{(j)}(\lambda_i)(\lambda-\lambda_i)^j}{j!}}{(\lambda-\lambda_i)^{l_i}}=\frac{p(\lambda)}{m_A(\lambda)}-\sum_{j=0}^{l_i-1}\frac{p_i^{(j)}(\lambda_i)(\lambda-\lambda_i)^{j-l_i}}{j!}$$

在 $\lambda=\lambda_i$ 的空心邻域内解析,进而有

$$\frac{p(\lambda)}{m_A(\lambda)}-\sum_{i=1}^{t}\sum_{j=0}^{l_i-1}\frac{p_i^{(j)}(\lambda_i)(\lambda-\lambda_i)^{j-l_i}}{j!}\triangleq g(\lambda)$$

在所有可能的可去奇点 $\lambda_1,\cdots,\lambda_t$ 外解析,从而适当补充 $g(\lambda)$ 在 $\lambda_1,\cdots,\lambda_t$ 各点的值后 $g(\lambda)$ 为全纯函数.再由$\lim\limits_{\lambda\to\infty}g(\lambda)=0$,因此利用刘维尔定理知 $g(\lambda)\equiv0$,即有

$$\frac{p(\lambda)}{m_A(\lambda)}=\sum_{i=1}^{t}\sum_{j=0}^{l_i-1}\frac{p_i^{(j)}(\lambda_i)(\lambda-\lambda_i)^{j-l_i}}{j!}$$

□

注意:

(1)若 $A=PBP^{-1}$,则任意在 Λ_A 上确定的复函数$f(\lambda)$在 Λ_B 上也确定,因为 A 的谱与 B 的谱完全一致.因此,$f(A)$的定义多项式 $p(\lambda)$也可充当$f(B)$的定义多项式,结果$f(A)$

$= Pf(B)P^{-1}$.

(2)$f(A_1 \oplus \cdots \oplus A_k) = f(A_1) \oplus \cdots \oplus f(A_k)$.事实上,$A = A_1 \oplus \cdots \oplus A_k$ 的最小多项式 $m_A(\lambda)$更是每个 A_i 的化零多项式.当 f 在 Λ_A 上确定时,f 在每个 Λ_{A_i} 上也确定,从而$f(A)$的任一定义多项式 $p(\lambda)$也可作为 $f(A_i)$的定义多项式.由多项式的性质知:

$$f(A_1 \oplus \cdots \oplus A_k) = p(A_1 \oplus \cdots \oplus A_k) = p(A_1) \oplus \cdots \oplus p(A_k) = f(A_1) \oplus \cdots \oplus f(A_k)$$

现在给定方阵 A,让 A 的 Jordan 分解为 $A = S(J_1 \oplus \cdots \oplus J_m)S^{-1}$,其中

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{l_i \times l_i}$$

若 $f(\lambda)$为在 Λ_A 上确定的复函数,于是基于上述(1)、(2)两点,得

$$f(A) = S(f(J_1) \oplus \cdots \oplus f(J_m))S^{-1} \tag{5.1.2}$$

考察 $f(J_i)$,J_i 的最小多项式为$(\lambda - \lambda_i)^{l_i}$,令多项式 $p(\lambda)$为

$$p(\lambda) = f(\lambda_i) + \frac{f'(\lambda_i)}{1!}(\lambda - \lambda_i) + \cdots + \frac{f^{(l_i-1)}(\lambda_i)}{(l_i-1)!}(\lambda - \lambda_i)^{l_i-1}$$

则直接验证知:

$$\begin{cases} p(\Lambda_{J_i}) = f(\Lambda_{J_i}) \\ f(J_i) = p(J_i) = \begin{pmatrix} f(\lambda_i) & \frac{f'(\lambda_i)}{1!} & \cdots & \frac{f^{(l_i-1)}(\lambda_i)}{(l_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{pmatrix} \end{cases} \tag{5.1.3}$$

因此,通过矩阵的 Jordan 标准形,按(5.1.2)式和(5.1.3)式也可给出 $f(A)$的定义,它与定义 5.1.3 是等价的.

对于性态较为良好的复函数 $f(z)$,矩阵函数 $f(A)$的定义还可通过矩阵级数的收敛极限来给出,其具体过程详见下述定理.

定理 5.1.3 设复变函数 $f(z)$在开圆域$|z - z_0| < r$ 内解析,即 $f(z)$在此开圆域内可展开成幂级数

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$$

若方阵 A 的所有特征值均在此开圆域内,则有

$$f(A) = \sum_{k=0}^{\infty} (A - z_0 I)^k$$

式中 $f(A)$ 为定义 5.1.3 中所定义的矩阵函数值.

证明: A 的所有特征值均在收敛圆盘$|z - z_0| < r$ 内,因此$\sum\limits_{k=0}^{\infty}(A - z_0 I)^k$ 收敛.记 B

$= \sum_{k=0}^{\infty}(A - z_0 I)^k$，下证 $f(A) = B$.先将 A 化为 Jordan 标准形

$$A = S(J_1 \oplus \cdots \oplus J_m)S^{-1}$$

于是

$$f(A) = S(f(J_1) \oplus \cdots \oplus f(J_m))S^{-1}$$

式中

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{l_i \times l_i}$$

$$f(J_i) = \begin{pmatrix} f(\lambda_i) & \dfrac{f'(\lambda_i)}{1!} & \cdots & \dfrac{f^{(l_i-1)}(\lambda_i)}{(l_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \dfrac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{pmatrix}$$

注意到

$$(J_i - z_0 I)^k = \begin{pmatrix} \mu_i^k & C_k^1 \mu_i^{k-1} & \cdots & C_k^{l_i-1}\mu_i^{k-l_i+1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & C_k^1 \mu_i^{k-1} \\ & & & \mu_i^k \end{pmatrix}$$

式中 $\mu_i = \lambda_i - z_0$. 当 $j > k$ 时约定 $C_k^j = 0$.而由 $f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k$ 又知

$$\begin{cases} f(\lambda_i) = \sum_{k=0}^{\infty} a_k \mu_i^k \\ \dfrac{f'(\lambda_i)}{1!} = \sum_{k=1}^{\infty} a_k C_k^1 \mu_i^{k-1} \\ \vdots \\ \dfrac{f^{(l_i-1)}(\lambda_i)}{(l_i-1)!} = \sum_{k=l_i-1}^{\infty} a_k C_k^{l_i-1} \mu_i^{k-l_i+1} \end{cases}$$

直接验算知：

$$\sum_{k=0}^{\infty} a_k (J_i - z_0 I)^k = f(J_i)$$

由此

$$\sum_{k=0}^{\infty} a_k (A - z_0 I)^k = S\Big(\sum_{k=0}^{\infty} a_k [(J_1 - z_0 I) \oplus \cdots \oplus (J_m - z_0 I)^k]\Big)S^{-1}$$

$$= S\Big(\sum_{k=0}^{\infty} a_k(J_1 - z_0 I)^k \oplus \cdots \oplus \sum_{k=0}^{\infty} a_k(J_m - z_0 I)^k\Big)S^{-1}$$
$$= S(f(J_1) \oplus \cdots \oplus f(J_m))S^{-1}$$
$$= f(A)$$

□

值得指出的是，当给定的矩阵 A 是较特殊的矩阵，例如规范阵时，按定义 5.1.3 给出的矩阵函数的定义又可有以下等价形式：

对于$\mathbb{C}$ 中的规范阵 A，取 A 的谱分解：$A = \lambda_1 E_1 + \cdots + \lambda_k E_k$，其中 $\lambda_1, \cdots, \lambda_k$ 为A 的所有互异特征值，$E_1, \cdots, E_k$ 为非零的两两正交的垂直射影，$E_1 + \cdots + E_k = 1$.$f(z)$ 为在 $\lambda_1, \cdots, \lambda_k$ 上均有定义的复函数，则 $f(A)$ 可定义为：

$$f(A) = f(\lambda_1)E_1 + \cdots + f(\lambda_k)E_k$$

为说明 $f(\lambda_1)E_1 + \cdots + f(\lambda_k)E_k$ 即为按定义 5.1.3 给出的矩阵函数值 $f(A)$，我们通过以下两步来进行.

首先，A 的规范性导数 A 的最小多项式 $m_A(\lambda)$为$(\lambda - \lambda_1)\cdots(\lambda - \lambda_k)$，从而 $f(z)$在 Λ_A 上确定.

其次，按照定义 5.1.3，$f(A)$的定义多项式 $p(\lambda)$可直接取为

$$p(\lambda) = f(\lambda_1)p_1(\lambda) + \cdots + f(\lambda_k)p_k(\lambda)$$

式中

$$p_i(\lambda) = \frac{(\lambda - \lambda_1)\cdots(\lambda - \lambda_{i-1})(\lambda - \lambda_{i+1})\cdots(\lambda - \lambda_k)}{(\lambda_i - \lambda_1)\cdots(\lambda_i - \lambda_{i-1})(\lambda_i - \lambda_{i+1})\cdots(\lambda_i - \lambda_k)}, i = 1,2,\cdots,k$$

事实上，$\{p_i(\lambda)\}_{i=1}^k$具备下列性质：

(1)$p_i(\lambda_i) = 1, p_i(\lambda_j) = 0(j \neq i), i,j = 1,2,\cdots,k$;

(2)$p_i(A) = p_i(\lambda_1 E_1 + \cdots + \lambda_k E_k) = p_i(\lambda_1)E_1 + \cdots + p_i(\lambda_k)E_k = E_i$.

因此

$$\begin{cases} p(\Lambda_A) = f(\Lambda_A) \\ f(A) = p(A) \end{cases}$$

亦即

$$f(A) = f(\lambda_1)E_1 + \cdots + f(\lambda_k)E_k$$

□

5.2　矩阵函数的性质及其初等因子

上一节解决了矩阵函数的定义问题.本质上说，它是矩阵多项式的推广.本节我们讨论矩阵函数的性质及其初等因子问题.

定理 5.2.1　设 $G(y_1, \cdots, y_l)$是 l 元多项式，函数 $f_1(\lambda), \cdots, f_l(\lambda)$都在 Λ_A 上确定.记 $g(\lambda) = G(f_1(\lambda), \cdots, f_l(\lambda))$，则有 $g(A) = G(f_1(A), \cdots, f_l(A))$.特别地，若 $g(\Lambda_A) = 0$，则

有 $G(f_1(A),\cdots,f_l(A))=0$.

证明: 取 $p_1(\lambda),\cdots,p_l(\lambda)$分别为 $f_1(A),\cdots,f_l(A)$的定义多项式,于是 $\varphi(\lambda)=G(p_1(\lambda),\cdots,p_l(\lambda))$仍为多项式,且有 $\varphi(\Lambda_A)=g(\Lambda_A)$,结果

$$g(A)=G(p_1(A),\cdots,p_l(A))=G(f_1(A),\cdots,f_l(A))$$

当 $g(\Lambda_A)=0$ 时,由 $g(A)=0$ 知,$G(f_1(A),\cdots,f_l(A))=0$.

□

定理 5.2.2 设复合函数 $g(\lambda)=h(f(\lambda))$在 Λ_A 上确定,则有

$$g(A)=h(f(A))$$

证明: 记 A 的最小多项式为 $m_A(\lambda)=(\lambda-\lambda_1)^{l_1}\cdots(\lambda-\lambda_t)^{l_t}$, $B=f(A)$. 又记

$$\begin{cases} f(\lambda_1)=\mu_1 \\ \quad\vdots \\ f(\lambda_t)=\mu_t \end{cases}$$

考虑多项式 $\varphi(\mu)=(\mu-\mu_1)^{l_1}\cdots(\mu-\mu_t)^{l_t}$,可设 $\mu_1,\cdots,\mu_t$ 各不相同. 若 $\mu_1=\mu_2$, $l_1\geq l_2$,则 $\varphi(\mu)$中只有$(\mu-\mu_1)l_1$ 含 $\mu-\mu_1$ 因子 φ.由于函数

$$\psi(\lambda)\triangleq\varphi(f(\lambda))=(f(\lambda)-\mu_1)^{l_1}\cdots(f(\lambda)-\mu_t)^{l_t}$$

满足 $\psi(\Lambda_A)=0$.因此由定理 5.2.1 得 $\psi(A)=\varphi(B)=0$.结果 $\varphi(\mu)$为 B 的零化多项式,进而 B 的最小多项式为

$$m_B(\mu)=(\mu-\mu_1)^{m_1}\cdots(\mu-\mu_t)^{m_t}$$

式中 $m_i\leq l_i$, $i=1,2,\cdots,t$.对函数 $h(\mu)$,先取最小多项式为$(\mu-\mu_1)^{l_1}\cdots(\mu-\mu_t)^{l_t}$的矩阵$C$,再取多项式 $r(\mu)$为 $h(C)$的定义多项式,从而 $r(\mu)$又可充当 $h(B)$的定义多项式,$r(B)=h(B)$, $r^{(j)}(\mu_i)=h^{(j)}(\mu_i)$, $i=1,2,\cdots,t$, $j=0,1,\cdots,l_i-1$.

比较函数 $g(\lambda)=h(f(\lambda))$和 $w(\lambda)=r(f(\lambda))$,由复合函数的求导法则知:

$$\begin{cases} g(\lambda_i)=w(\lambda_i) \\ g'(\lambda_i)=w'(\lambda_i) \\ \quad\vdots \\ g^{(l_i-1)}(\lambda_i)=w^{(l_i-1)}(\lambda_i) \end{cases} \qquad i=1,2,\cdots,t$$

结果$(g(\lambda)-w(\lambda))(\Lambda_A)=0$,注意到 $g(\lambda)-w(\lambda)$是关于 $g(\lambda)$与 $f(\lambda)$的二元多项式,因此利用定理 5.2.1 得 $g(A)-r(f(A))=0$,即

$$g(A)=r(f(A))=h(f(A))$$

□

定理 5.2.1 与定理 5.2.2 是很有用的,许多矩阵函数的等式问题都可通过它们来解决.

例 5.2.1 (1)由 $\cos^2x+\sin^2x=1$ 知,对任何 n 阶方阵A 均有

$$\cos^2A+\sin^2A=I$$

(2)$\forall A,B\in\mathbb{C}^{n\times n}$, $k,l\in\mathbb{C}$,由 $\mathrm{e}^{kz}\mathrm{e}^{lz}=\mathrm{e}^{(k+l)z}$知

$$\mathrm{e}^{(k+l)A}=\mathrm{e}^{kA}\mathrm{e}^{lA}$$

$$(e^{A})^{-1}=e^{-A}$$

$$e^{A+2\pi iI}=e^{A}$$

当 $AB=BA$ 时,由 $\sum\limits_{k=0}^{\infty}\frac{A^k}{k!}$ 和 $\sum\limits_{k=0}^{\infty}\frac{B^k}{k!}$ 均绝对收敛知,$\sum\limits_{k=0}^{\infty}\frac{A^k}{k!}\sum\limits_{k=0}^{\infty}\frac{B^k}{k!}$ 按任何方式排列得到的级数也绝对收敛,且其和均为 e^Ae^B.特别,按如下方式排列 $\sum\limits_{k=0}^{\infty}\frac{A^k}{k!}\sum\limits_{k=0}^{\infty}\frac{B^k}{k!}$:

$$\sum_{k=0}^{\infty}C_k=I\cdot I+\frac{(A+B)}{1!}+\cdots+\frac{(A+B)^m}{m!}+\cdots$$

得:$e^{A+B}=e^Ae^B=e^Be^A$.

□

例 5.2.2　设 A 的 Jordan 分解为

$$A=S(J_1\oplus\cdots\oplus J_t)S^{-1}$$

$$J_i=\begin{pmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda_i\end{pmatrix}_{l_i\times l_i}$$

于是

$$e^A=S(e^{J_1}\oplus\cdots\oplus e^{J_t})S^{-1},e^{J_i}=\begin{pmatrix}e^{\lambda_i} & \cdots & \frac{e^{\lambda_i}}{(l_i-1)!}\\ & \ddots & \vdots\\ & & e^{\lambda_i}\end{pmatrix}$$

因此 $\det e^A=e^{\operatorname{tr}A}\neq 0$,$e^A$ 总是可逆的.

另一方面,若 A 为可逆方阵,则 A 的所有复特征根均非零,从而在复平面一定有以 A 的所有特征根为内点,但又不含原点的单连通区域.在这个区域内可以确定一个单值解析函数 $\ln z$ 使得 $z=e^{\ln z}$ 成立.由定理 5.2.2 可在 $A=e^{\ln A}$.

□

例 5.2.2 表明,下列映照

$$\sigma:(\mathbb{C}^{n\times n},+)\to(GL(n,\mathbb{C}),\cdot)$$

$$A\to e^A$$

为一个满映照.但值得注意 $\sigma(A+B)=\sigma(A)\sigma(B)$ 一般未必成立,从而 σ 不构成群同态.

例 5.2.3　让 $A=\begin{pmatrix}1&0\\0&0\end{pmatrix}$,$B=\begin{pmatrix}1&1\\0&0\end{pmatrix}$,于是

$$e^A=\begin{pmatrix}e&0\\0&0\end{pmatrix},\ e^B=\begin{pmatrix}e&e-1\\0&1\end{pmatrix}$$

$$A+B=\begin{pmatrix}2&1\\0&0\end{pmatrix},\ e^{A+B}=\begin{pmatrix}e^2&\frac{1}{2}(e^2-1)\\0&1\end{pmatrix}$$

$$e^A e^B = \begin{pmatrix} e^2 & e^2 - e \\ 0 & 0 \end{pmatrix}, \quad e^B e^A = \begin{pmatrix} e^2 & 0 \\ 0 & 0 \end{pmatrix}$$

从而 e^{A+B},$e^A e^B$,$e^B e^A$ 各不相同.

□

例 5.2.4 给定可逆方阵 A,矩阵方程 $X^p = A$ 对任何非零复数 p 均有解.

证明: A 可逆说明 A 的特征根均非零.从而复平面内存在一个包含 A 的所有特征根为内点,且不包含零的单连通区域,在此单连通区域上可以确定一个单值解析函数 $z^{\frac{1}{p}}$ 使得 $(z^{\frac{1}{p}})^p = z$.由定理 5.2.2 知 $(A^{\frac{1}{p}})^p = A$.

□

下面我们研究 $f(A)$ 的初等因子问题.设复方阵 A 的 Jordan 分解为 $A = S(J_1 \oplus \cdots \oplus J_m)S^{-1}$,其中 $J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$.当复函数 $f(z)$ 在 Λ_A 上确定时,亦即 $f(A)$ 有意义情况下,前面已得

$$f(J_i) = \begin{pmatrix} f(\lambda_i) & \dfrac{f'(\lambda_i)}{1!} & \cdots & \dfrac{f^{(l_i-1)}(\lambda_i)}{(l_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \dfrac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{pmatrix}$$

$$f(A) = S(f(J_1) \oplus \cdots \oplus f(J_m))S^{-1}$$

因此,确定 $f(A)$ 的初等因子可归结为求 $f(J_i)$ 的初等因子.由 $f(J_i)$ 的特殊形状,一般我们有:

命题 5.2.1 设 p 阶上三角方阵

$$B = \begin{pmatrix} a_0 & a_1 & \cdots & a_{p-1} \\ & \ddots & \ddots & \vdots \\ & & \ddots & a_1 \\ & & & a_0 \end{pmatrix}$$

则 (1) $a_1 \neq 0$ 时,B 的初等因子只有 $(\lambda - a_0)^p$;

(2) $a_1 = \cdots = a_{k-1} = 0$, $a_k \neq 0$ 时,B 的初等因子由 $(k-h)$ 个 $(\lambda - a_0)^q$ 和 h 个 $(\lambda - a_0)^{q+1}$ 组成,其中 q,h 由下列带余除法确定:$p = qk + h$,$(0 \leqslant h < k)$;

(3) $a_1 = \cdots = a_{p-1} = 0$ 时,B 的初等因子由 p 个 $(\lambda - a_0)$ 组成.

证明: (1) $a_1 \neq 0$ 时,秩$(B - a_0 I) = p - 1$,因此 $\lambda I - B$ 的各级行列式因子为

$$D_p(\lambda) = (\lambda - a_0)^p, D_{p-1}(\lambda) = 1, \cdots, D_1(\lambda) = 1, D_0(\lambda) = 1,$$

进而 $\lambda I - B$ 的不变因子组为

$$(\lambda - a_0)^p, 1, \cdots$$

从而 B 的初等因子只有$(\lambda - a_0)^p$.

(2)$a_1 = \cdots = a_{k-1} = 0, a_k \neq 0$ 情形下,有

$$B = a_0 I + a_k H^k + \cdots + a_{p-1} H^{p-1}$$

式中 $H = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$.此时有

$$(B - a_0 I)^j = a_k^j H^{kj} + \text{若干形如 } cH^l \text{ 的项}$$

其中幂次 l 比 kj 更大.所以

$$\text{秩}(B - a_0 I)^j = \begin{cases} 0, & kj > p \\ p - kj, & kj \leqslant p \end{cases}$$

令 $p = qk + h, (0 \leqslant h < k)$,则 B 关于 a_0 而阶为 q 的 Jordan 块个数为:

$$\begin{aligned} &\text{秩}(B - a_0 I)^{q+1} + \text{秩}(B - a_0 I)^{q-1} - 2\,\text{秩}(B - a_0 I)^q \\ &= 0 + [p - (kq - k)] - 2(p - kq) \\ &= k - h \end{aligned}$$

同样,B 关于 a_0 阶为 $q+1$ 的 Jordan 块个数为:

$$\text{秩}(B - a_0 I)^{q+2} + \text{秩}(B - a_0 I)^q - 2\,\text{秩}(B - a_0 I)^{q+1} = 0 + (p - kq) - 2\cdot 0 = h$$

显然,由$(k-h)q + h(q+1) = p$ 知:B 除了 $k-h$ 个$(\lambda - a_0)^q$ 及 h 个$(\lambda - a_0)^{q+1}$外并无其他的初等因子.

(3)是显然的.

□

通过命题 5.2.1,$f(A)$的全部初等因子问题便可按下述定理 5.2.3 中的方式完全给出.

定理 5.2.3 设 J 为 n 阶方阵 A 的一个 Jordan 块,J 对应的初等因子为$(\lambda - \lambda_0)^p$,则 $f(J)$的初等因子为:

(1)当 $p = 1$ 时,$f(J)$的初等因子为$(\lambda - f(\lambda_0))$;

(2)当 $p > 1$ 但 $f'(\lambda_0) \neq 0$ 时,$f(J)$的初等因子只有$(\lambda - f(\lambda_0))^p$;

(3)当 $p > 1, f^{(1)}(\lambda_0) = \cdots = f^{(k-1)}(\lambda_0) = 0, f^{(k)}(\lambda_0) \neq 0$ 时$(k \leqslant p-1)$,$f(J)$的初等因子由 $k-h$ 个$(\lambda - f(\lambda_0))^q$ 及 h 个$(\lambda - f(\lambda_0))^{q+1}$组成,其中 q, h 由下列带余除法确定:$p = qk + h, (0 \leqslant h < k)$;

(4)当 $p > 1$,但 $f^{(1)}(\lambda_0) = \cdots = f^{(p-1)}(\lambda_0) = 0$ 时,$f(J)$的初等因子由 p 个$(\lambda - f(\lambda_0))$组成.

证明: 由命题 5.2.1,显然.

□

推论 5.2.1 若对 A 的每个特征值 λ_i 均有 $f'(\lambda_i) \neq 0$,则 A 的每个初等因子$(\lambda -$

$\lambda_i)^{m_{ij}}$对应着$f(A)$的一个初等因子$(\lambda - f(\lambda_i))^{m_{ij}}$. 此时,记 A 的 Jordan 标准形为

$$\begin{pmatrix} \lambda_1 & 1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & \ddots & 1 & & & & & \\ & & & \lambda_1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \lambda_t & 1 & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & \ddots & 1 \\ & & & & & & & & \lambda_t \end{pmatrix}$$

则 $f(A)$的 Jordan 标准形即为

$$\begin{pmatrix} f(\lambda_1) & 1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & \ddots & 1 & & & & & \\ & & & f(\lambda_1) & & & & & \\ & & & & \ddots & & & & \\ & & & & & f(\lambda_t) & 1 & & \\ & & & & & & \ddots & \ddots & \\ & & & & & & & \ddots & 1 \\ & & & & & & & & f(\lambda_t) \end{pmatrix}$$

5.3 矩阵函数的计算

矩阵函数的计算问题是矩阵应用中常遇到的关键问题之一.矩阵函数的计算通常是相当复杂的,因此研究方便快捷的计算矩阵函数的相关方法成为人们感兴趣的话题.本节主要介绍几种常用的计算矩阵函数的方法.

(1)递推公式计算法

该方法的主要原理是基于 Caley - Hamilton 定理,A 的特征多项式 $\det(\lambda I - A)$即为 A 的一个化零多项式,从而由多项式的带余除法,$A^j\,(j \geq n)$总可表为关于 $A^{n-1},\cdots,A$ 的一个多项式.由此得到 A 的递推关系式,进而利用函数的幂级数展式计算出矩阵 A 的函数值.**此方法中关键是利用特征多项式,由多项式的带余除法得到 A 的递推关系式**.

例 5.3.1 设 4 阶方阵 A 的特征值为 $\mathrm{e}, -\mathrm{e}, 0, 0$,求 $\sin A$.

解: A 的特征多项式为$(\lambda - \mathrm{e})(\lambda + \mathrm{e})\lambda^2 = \lambda^4 - \lambda^2\mathrm{e}^2$.因此得 A 的递推公式:

$$A^4 = \mathrm{e}^2 A^2$$

一般地,

$$A^5 = AA^4 = \mathrm{e}^2 A^3 = \mathrm{e}^{5-3} A^3$$

$$A^7 = A^2A^5 = e^2A^5 = e^{7-3}A^3$$
$$\vdots$$
$$A^{2k+1} = e^{(2k+1)-3}A^3$$
$$\vdots$$

结果：

$$\begin{aligned}\sin A &= A - \frac{A^3}{3!} + \frac{A^5}{5!} + \cdots + (-1)^k \frac{A^{2k+1}}{(2k+1)!} + \cdots \\ &= A - \frac{A^3}{3!} + \frac{e^2A^3}{5!} + \cdots + (-1)^k \frac{e^{(2k+1)-3}A^3}{(2k+1)!} + \cdots \\ &= A + \frac{A^3}{e^3}[-e + \sin e]\end{aligned}$$

□

(2)基于 Jordan 标准形的算法

设 $A = S(J_1 \oplus \cdots \oplus J_m)S^{-1}$，其中

$$J_i = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}_{l_i \times l_i}$$

前面已证明：$f(A) = S(f(J_1) \oplus \cdots \oplus f(J_m))S^{-1}$，其中

$$f(J_i) = \begin{pmatrix} f(\lambda_i) & \frac{f'(\lambda_i)}{1!} & \cdots & \frac{f^{(l_i-1)}(\lambda_i)}{(l_i-1)!} \\ & \ddots & \ddots & \vdots \\ & & \ddots & \frac{f'(\lambda_i)}{1!} \\ & & & f(\lambda_i) \end{pmatrix}$$

因此，**此方法的关键是求 A 的 Jordan 标准形及其变换阵 S.**

例 5.3.2　设

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & -1 & 3 \end{pmatrix}$$

求 A^{20}，e^A，$\sin A$，e^{At}.

解： 首先，利用 λ 阵，得到 A 的 Jordan 标准形 $J = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$，再由 $AP = PJ$ 可求得变换阵 $P = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$，$P^{-1} = \begin{pmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}$.

其次，记 $f_1(z) = z^{20}$，$f_2(z) = e^z$，$f_3(z) = \sin z$，$f_4(z) = e^{zt}$，于是 $f'_1(2) = 20 \cdot 2^{19}$，$f'_2(2)$

$=\mathrm{e}^2, f'_3(2)=\cos 2, f'_4(2)=t\mathrm{e}^{2t}$.结果

$$f_1(J)=\begin{pmatrix} f_1(2) & 0 & 0 \\ 0 & f_1(2) & f'_1(2) \\ 0 & 0 & f_1(2) \end{pmatrix}=2^{20}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 10 \\ 0 & 0 & 1 \end{pmatrix}$$

$$f_2(J)=\begin{pmatrix} \mathrm{e}^2 & 0 & 0 \\ 0 & \mathrm{e}^2 & \mathrm{e}^2 \\ 0 & 0 & \mathrm{e}^2 \end{pmatrix}=\mathrm{e}^2\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$f_3(J)=\begin{pmatrix} \sin 2 & 0 & 0 \\ 0 & \sin 2 & \cos 2 \\ 0 & 0 & \sin 2 \end{pmatrix}$$

$$f_4(J)=\begin{pmatrix} \mathrm{e}^{2t} & 0 & 0 \\ 0 & \mathrm{e}^{2t} & t\mathrm{e}^{2t} \\ 0 & 0 & \mathrm{e}^{2t} \end{pmatrix}$$

最后

$$f_1(A)=Pf_1(J)P^{-1}=2^{20}\begin{pmatrix} 1 & 0 & 0 \\ 10 & -9 & 10 \\ 10 & -10 & 11 \end{pmatrix}$$

$$f_2(A)=Pf_2(J)P^{-1}=\mathrm{e}^2\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 2 \end{pmatrix}$$

$$f_3(A)=Pf_3(J)P^{-1}=\sin 2\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}+\cos 2\begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{pmatrix}$$

$$f_4(A)=Pf_4(J)P^{-1}=\mathrm{e}^{2t}\begin{pmatrix} 1 & 0 & 0 \\ t & 1-t & t \\ t & -t & 1+t \end{pmatrix}$$

(3)拉格朗日插值法

此方法的根据在于定理5.1.2.记给定的矩阵 A 的最小多项式为 $m_A(\lambda)=(\lambda-\lambda_1)^{l_1}\cdots(\lambda-\lambda_t)^{l_t}$,$f(\lambda)$在 Λ_A 上确定,则 $f(A)=p(A)$,其中 $p(\lambda)$为多项式

$$p(\lambda)=m_A(\lambda)\sum_{i=1}^{t}\left[\frac{f_i(\lambda_i)}{(\lambda-\lambda_i)^{l_i}}+\frac{f_i^{(1)}(\lambda_i)/1!}{(\lambda-\lambda_i)^{l_i-1}}+\cdots+\frac{f_i^{(l_i-1)}(\lambda_i)/(l_i-1)!}{\lambda-\lambda_i}\right] \tag{5.3.1}$$

式中 $f_i(\lambda)=\dfrac{f(\lambda)(\lambda-\lambda_i)^{l_i}}{m_A(\lambda)}$.因此,**拉格朗日插值法的关键是求 $f_i(\lambda)$在 $\lambda=\lambda_i$ 处的各阶导数值**.

回顾定理5.1.2中(2)的证明过程中,主要使用的是复变函数论中的技巧.下面的推导表明该论断还可利用初等的方法加以证明.

首先，将其分式$\frac{p(\lambda)}{m_A(\lambda)}$分解为

$$\frac{p(\lambda)}{m_A(\lambda)} = \sum_{i=1}^{t}\left[\frac{a_{i0}}{(\lambda-\lambda_i)^{l_i}} + \frac{a_{i1}}{(\lambda-\lambda_i)^{l_i-1}} + \cdots + \frac{a_{i(l_i-1)}}{\lambda-\lambda_i}\right]$$

亦即

$$(\lambda-\lambda_i)^{l_i}\frac{p(\lambda)}{m_A(\lambda)} = a_{i0} + a_{i1}(\lambda-\lambda_i) + \cdots + a_{i(l_i-1)}(\lambda-\lambda_i)^{l_i-1} + (\lambda-\lambda_i)^{l_i}\sum_{k\neq i}(\cdots) \tag{5.3.2}$$

其次，为求 a_{i0}，上式两边让 $\lambda\to\lambda_i$ 得

$$a_{i0} = f(\lambda_i)\left(\left.\frac{(\lambda-\lambda_i)^{l_i}}{m_A(\lambda)}\right|_{\lambda=\lambda_i}\right) = f_i(\lambda_i)$$

为求 a_{i1}，将式(5.3.2)两边求导，再让 $\lambda\to\lambda_i$ 得

$$a_{i1} = \left.\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[(\lambda-\lambda_i)^{l_i}\frac{p(\lambda)}{m_A(\lambda)}\right]\right|_{\lambda=\lambda_i} = \frac{f_i^{(1)}(\lambda_i)}{1!}$$

类似地，有

$$a_{i(l_i-1)} = \frac{\left.\frac{\mathrm{d}^{(l_i-1)}}{\mathrm{d}\lambda^{(l_i-1)}}\left[(\lambda-\lambda_i)^{l_i}\frac{p(\lambda)}{m_A(\lambda)}\right]\right|_{\lambda=\lambda_i}}{(l_i-1)!} = \frac{f_i^{(l_i-1)}(\lambda_i)}{(l_i-1)!}$$

定理 5.1.2 中的(2)获证.

□

观察式(5.3.1)会发现，若归并 $f(\lambda)$在各 λ_i 处的各阶导数后，$f(A)$的定义多项式 $p(\lambda)$可以写成

$$p(\lambda) = \sum_{i=1}^{t}\left[f_i(\lambda_i)\varphi_{i1}(\lambda) + f_i^{(1)}(\lambda_i)\varphi_{i2}(\lambda) + \cdots + f_i^{(l_i-1)}(\lambda_i)\varphi_{il_i}(\lambda)\right]$$

其中各 $\varphi_{ij}(\lambda)$ 均满足 $\deg\varphi_{ij}(\lambda) < \deg m_A(\lambda)$，且与 $f(\lambda)$ 无关，仅由 $m_A(\lambda)$ 确定.因此

$$f(A) = p(A) = \sum_{i=1}^{t}\left[f_i(\lambda_i)\varphi_{i1}(A) + f_i^{(1)}(\lambda_i)\varphi_{i2}(A) + \cdots + f_i^{(l_i-1)}(\lambda_i)\varphi_{il_i}(A)\right]$$

其中 $\varphi_{ij}(A)$仅由 A 所确定，与 $f(\lambda)$无关.这一性质有时对我们计算 $f(A)$能带来很大方便.

例 5.3.3　让 $A = \begin{pmatrix} 0 & -1 \\ 4 & 4 \end{pmatrix}$，求 $\arcsin\left(\frac{A}{4}\right)$.

解： 令 $f(z) = \arcsin\left(\frac{z}{4}\right)$，$\det(\lambda I - A) = (\lambda-2)^2$，$m_A(\lambda) = (\lambda-2)^2$.因此可令

$$f(A) = f(2)\varphi_{11}(A) + f'(2)\varphi_{12}(A) = \frac{\pi}{6}\varphi_{11}(A) + \frac{\sqrt{3}}{6}\varphi_{12}(A)$$

为求 $\varphi_{11}(A)$与 $\varphi_{12}(A)$，我们令 $g(z) = z-2$，于是有

$$A - 2I = \varphi_{12}(A) = \begin{pmatrix} -2 & -1 \\ 4 & 2 \end{pmatrix}$$

再令 $g(z)=1$,得

$$I=\varphi_{11}(A)$$

结果
$$f(A)=\frac{\pi}{6}I+\frac{\sqrt{3}}{6}(A-2I)=\frac{1}{6}\begin{pmatrix}\pi-2\sqrt{3} & -\sqrt{3}\\ 4\sqrt{3} & \pi+2\sqrt{3}\end{pmatrix}$$

□

(4)待定系数法

此方法的依据是定理 5.1.2 和定义 5.1.3.设矩阵 A 的最小多项式的次数是 m.于是可令 $f(A)$的定义多项式 $p(x)$为

$$p(x)=a_0+a_1x+\cdots+a_{m-1}x^{m-1}$$

其中 $a_0,\cdots,a_{m-1}$待定.再由定义 5.1.3 得线性方程组

$$p(\Lambda_A)=f(\Lambda_A)$$

由该线性方程组求得 $a_0,\cdots,a_{m-1}$,从而 $f(A)=p(A)$.明显地,若 m 较小时,此种方法具有较大的优越性.

习题五

1. 设 $A\in\mathbb{C}^{n\times n}$, $\|A\|_2<1$.证明 $\|\ln(I_n+A)\|_2\leqslant\frac{\|A\|_2}{1-\|A\|_2}$.

2. 证明:$\det(e^{At})=e^{(\mathrm{tr}A)t}$.

3. 设 $A\in\mathbb{C}^{n\times n}$,证明$(e^{At})^{\mathrm{T}}=e^{A^{\mathrm{T}}t}$.

4. 计算级数$\sum\limits_{n=0}^{\infty}\frac{1}{10^n}\begin{pmatrix}1 & 2\\ 8 & 1\end{pmatrix}^{n+1}$.

5. 证明:如果 A 是 Hermite 矩阵,则 e^{iA}为酉阵;如果 A 是实斜对称矩阵,则 e^{A} 为正交阵.

6. 设 $A\in\mathbb{C}^{n\times n}$,$A$ 为可逆矩阵.试用 A 的多项式来表示 A 的逆矩阵.

第6章　函数矩阵的微积分及应用

本章的主要目标是,对元素为函数的矩阵建立相应的微积分公式.作为函数矩阵微积分的应用,重点介绍变系数线性微分方程组的求解.如无特别说明,本章涉及的函数均是指实函数.

6.1　基本概念及性质

元素为函数的矩阵,称为**函数矩阵**.例如 $A(t)=(a_{ij}(t))_{m\times n}$ 便是一个变量为 t 的函数矩阵.一般地,对函数矩阵 $A(t)$,若 $\forall\, i,j$ 均有 $\lim\limits_{t\to t_0}a_{ij}(t)=a_{ij}$,则称 **$A(t)$ 在 $t\to t_0$ 时极限为** $A=(a_{ij})$;若函数矩阵 $A(t)$ 中的所有元素 $a_{ij}(t)$ 在 t_0 处皆连续,则称 **$A(t)$ 在 t_0 处连续**.特别地,若 $A(t)$ 中每个元素 $a_{ij}(t)$ 在 (a,b) 内连续,则称 $A(t)$ 在 (a,b) 内连续.类似地,可定义 $A(t)$ 在 $[a,b]$ 上连续的概念.

函数矩阵的基本性质有:

性质 6.1.1　设 $\lim\limits_{t\to t_0}A(t)=A$, $\lim\limits_{t\to t_0}B(t)=B$,则

(1) $A(t)+B(t)$ 有定义时, $\lim\limits_{t\to t_0}(A(t)\pm B(t))=A\pm B$;

(2) $A(t)\cdot B(t)$ 有意义时, $\lim\limits_{t\to t_0}A(t)B(t)=AB$;

(3) $\lim\limits_{t\to t_0}kA(t)=kA$(其中 k 为常数).

定义 6.1.1　设函数矩阵 $A(t)$ 中所有元素 $a_{ij}(t)$ 均在 t_0 或在某区域内可微,则称函数矩阵 **$A(t)$ 在 t_0 处或某区域内是可微的**.当 $A(t)$ 可微时,定义 $A(t)$ 关于 t 的导数为:

$$A'(t)=(a'_{ij}(t))$$

类似地,可定义 $A(t)$ 的高阶导数的概念.

性质 6.1.2　设函数矩阵 $A(t)$, $B(t)$ 都可微,则有

(1) $[kA(t)]'=kA'(t)$,(k 为常数);

(2) 当 $A(t)+B(t)$ 有意义时, $[A(t)+B(t)]'=A'(t)+B'(t)$;

(3) 当 $A(t)B(t)$ 有意义时, $[A(t)B(t)]'=A'(t)B(t)+A(t)B'(t)$;

若 $A(t)$ 或 $B(t)$ 为常数矩阵 A 或 B 时, $[AB(t)]'=AB'(t)$, $[A(t)B]'=A'(t)B$;

(4) 设 A 为常数方阵,则 $(e^{At})'=Ae^{At}$; $(\sin At)'=A\cos At=(\cos At)A$;

$(\cos At)'=-A\sin At=-(\sin At)A$;

(5) $$\frac{\mathrm{d}}{\mathrm{d}t}\begin{vmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & \cdots & a_{2n}(t) \\ \vdots & & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{vmatrix} = \begin{vmatrix} a'_{11}(t) & \cdots & a'_{1n}(t) \\ a_{21}(t) & \cdots & a_{2n}(t) \\ \vdots & & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{vmatrix} + \cdots + \begin{vmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ a_{21}(t) & \cdots & a_{2n}(t) \\ \vdots & & \vdots \\ a'_{n1}(t) & \cdots & a'_{nn}(t) \end{vmatrix}$$

性质 6.1.3 (1)若 $A(w)$是关于 w 可微的函数矩阵,而 $w=f(t)$又是一元可微函数,则$[A(f(t))]'=A'_w(w)f'(t)$;

(2)若 $A(t)$及其逆矩阵 $A^{-1}(t)$均为可微函数,则有$[A^{-1}(t)]'=-A^{-1}(t)A'(t)A^{-1}(t)$.

证明: 只证(2),由 $A(t)\cdot A^{-1}(t)=I$,两边求导得

$$A'(t)A^{-1}(t)+A(t)[A^{-1}(t)]'=0$$

因此

$$[A^{-1}(t)]'=-A^{-1}(t)A'(t)A^{-1}(t)$$

□

例 6.1.1 设向量 $X=(x_1(t),\cdots,x_n(t))^{\mathrm{T}}$,$A(t)=(a_{ij}(t))_{n\times n}$为对称矩阵且皆是可微的,求$\frac{\mathrm{d}X^{\mathrm{T}}A(t)X}{\mathrm{d}t}$.

解:

$$\begin{aligned}(X^{\mathrm{T}}A(t)X)' &= (X^{\mathrm{T}})'A(t)X+X^{\mathrm{T}}(A(t)X)' \\ &= (X')^{\mathrm{T}}A(t)X+X^{\mathrm{T}}A'(t)X+X^{\mathrm{T}}A(t)X'\end{aligned}$$

由 $X^{\mathrm{T}}A(t)X'$为一阶矩阵,$X^{\mathrm{T}}A(t)X'=(X^{\mathrm{T}}A(t)X')^{\mathrm{T}}=(X')^{\mathrm{T}}A(t)X$,因此

$$\frac{\mathrm{d}X^{\mathrm{T}}A(t)X}{\mathrm{d}t}=X^{\mathrm{T}}A'(t)X+2X^{\mathrm{T}}A(t)X'$$

□

定义 6.1.2 若函数矩阵 $A(t)=(a_{ij}(t))_{m\times n}$ 中的每个元素 $a_{ij}(t)$ 在$[a,b]$上皆可积,则称 $A(t)$ 在$[a,b]$上可积,且定义$\int_a^b A(t)\mathrm{d}t=(\int_a^b a_{ij}(t)\mathrm{d}t)_{m\times n}$.

例如,设函数矩阵 $A(t)$ 为

$$A(t)=\begin{pmatrix}\sin t & -\cos t \\ \cos t & \sin t\end{pmatrix}$$

则$\int_0^{t^2}A(s)\mathrm{d}s=\begin{pmatrix}1-\cos t^2 & -\sin t^2 \\ \sin t^2 & 1-\cos t^2\end{pmatrix}$.

定理6.1.1 设 n 阶函数方阵$A(t)$在 $t\geqslant t_0$ 时连续,$x(t)$为 $n\times m$ 阶未知函数矩阵,则有

$$\begin{cases} \dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = A(t)x(t) \\ x(t_0) = C \end{cases}$$

在$[t_0, +\infty)$上存在唯一解,且该解可以写成

$$x(t) = X(t)C$$

的形式,式中 $X(t)$ 满足

$$\begin{cases} \dfrac{\mathrm{d}X(t)}{\mathrm{d}t} = A(t)X(t) \\ X(t_0) = I_n \end{cases} \tag{6.1.1}$$

证明: 易知,方程(6.1.1)存在解 $X(t)$ 等价于积分方程

$$X(t) = I_n + \int_{t_0}^{t} A(\tau)X(\tau)\mathrm{d}\tau$$

存在解 $X(t)$.为此,构造近似解序列:

$$\begin{cases} X_0 = I_n \\ X_1 = I_n + \displaystyle\int_{t_0}^{t} A(\tau)X_0\mathrm{d}\tau \\ \qquad\vdots \\ X_{k+1} = I_n + \displaystyle\int_{t_0}^{t} A(\tau)X_k\mathrm{d}\tau \\ \qquad\vdots \end{cases}$$

于是 $X_{k+1} - X_k = \int_{t_0}^{t} A(\tau)(X_k - X_{k-1})\mathrm{d}\tau$. $\forall\, t_1 > t_0$,由 $A(t)$ 在$[t_0, +\infty)$上连续,故可记 $m = \max\limits_{t\in[t_0,t_1]} \|A(t)\|_1$,其中 $\|A(t)\|_1 = \sum\limits_{i,j}^{n} |a_{ij}(t)|$.结果当 $t \in [t_0, t_1]$ 时,有

$$\begin{aligned} \|X_{k+1} - X_k\|_1 &\leqslant \int_{t_0}^{t} \|A(\tau)\|_1 \|X_k - X_{k-1}\|_1\mathrm{d}\tau \\ &\leqslant m\int_{t_0}^{t} \|X_k - X_{k-1}\|_1\mathrm{d}\tau \end{aligned}$$

于是由

$$\|X_1 - X_0\|_1 \leqslant \int_{t_0}^{t} \|A(\tau)\|_1\mathrm{d}\tau \leqslant m(t - t_0)$$

从而

$$\|X_{k+1} - X_k\|_1 \leqslant \frac{(m(t_1 - t_0))^{k+1}}{(k+1)!}$$

这表明矩阵级数 $\sum\limits_{k=0}^{\infty}(X_{k+1} - X_k)$ 在$[t_0, t_1]$上一致收敛,从而 X_k 一致收敛于 $x(t)$.该 $x(t)$ 连续, $x(t) = I_n + \int_0^t A(\tau)x(\tau)\mathrm{d}\tau$.

设 $y(t)$ 也是 $X(t) = I_n + \int_{t_0}^{t} A(\tau)X(\tau)\mathrm{d}\tau$ 的解,则有

$$\forall\, t \in [t_0, t_1], x(t) - y(t) = \int_{t_0}^{t} A(\tau)(x(\tau) - y(\tau))\mathrm{d}\tau \tag{6.1.2}$$

又由 $y(t)$ 在 $[t_0,t_1]$ 上连续，于是可记 $m_1=\max\limits_{t\in[t_0,t_1]}\|x(t)-y(t)\|_1$，此时有

$$\|x(t)-y(t)\|_1\leqslant m_1\int_{t_0}^{t}\|A(\tau)\|_1\mathrm{d}\tau$$

反复利用(6.1.2)式得：

$$\|x(t)-y(t)\|_1\leqslant m_1\frac{\left(\int_{t_0}^{t}\|A(\tau)\|_1\mathrm{d}\tau\right)^{k+1}}{(k+1)!}$$

由于 $\int_{t_0}^{t}\|A(\tau)\|_1\mathrm{d}\tau$ 在 $[t_0,t_1]$ 上是一个小于某个固定常数的数，因此有

$$\|x(t)-y(t)\|_1=0,\text{即 } x(t)=y(t),\text{对 } \forall t\in[t_0,t_1]$$

至此，由于 t_1 可任意大，故实际得到了当 $t\geqslant t_0$ 时的方程(6.1.1)的唯一解.对初始条件为 $x(t_0)=C$ 情形下的证明，由推导方程(6.1.1)解存在唯一性过程立即得知.

□

定义 6.1.3 若 $y=f(X)=f(x_{11},\cdots,x_{1n};x_{21},\cdots,x_{2n};\cdots;x_{m1},\cdots,x_{mn})$ 为对每个 x_{ij} 皆有偏导数的多元函数.定义数量函数 $y=f(X)$ 关于矩阵 $X=(x_{ij})_{m\times n}$ 导数为

$$\frac{\mathrm{d}f}{\mathrm{d}X}=\begin{pmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{pmatrix}$$

显见，当 $X=(x_1,\cdots,x_n)$ 时，$\frac{\mathrm{d}f}{\mathrm{d}X}=\left(\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_n}\right)$ 即为 f 的梯度 $\mathrm{grad}f$；而

$$\frac{\mathrm{d}f}{\mathrm{d}X^{\mathrm{T}}}=\begin{pmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{pmatrix}=\left(\frac{\mathrm{d}f}{\mathrm{d}X}\right)^{\mathrm{T}}$$

性质 6.1.4 设 $f(X),g(X)$ 为矩阵 X 的数量函数.若 $\frac{\mathrm{d}f}{\mathrm{d}X}$ 与 $\frac{\mathrm{d}g}{\mathrm{d}X}$ 皆存在，则

(1) $\frac{\mathrm{d}}{\mathrm{d}X}(f(x)\pm g(X))=\frac{\mathrm{d}f}{\mathrm{d}X}\pm\frac{\mathrm{d}g}{\mathrm{d}X}$；

(2) $\frac{\mathrm{d}}{\mathrm{d}X}(f(X)g(X))=f(X)\frac{\mathrm{d}g}{\mathrm{d}X}+\frac{\mathrm{d}f}{\mathrm{d}X}g(X)$；

(3) 让 $X=(x_{ij})_{m\times n},B=(b_{ij})_{n\times m},A=(a_{ij})_{m\times m}$ 有

$$\begin{cases}X^{\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}X}\det(X)=\det(X)I,X\text{ 为方阵时}\\ \frac{\mathrm{d}}{\mathrm{d}X}\mathrm{tr}(XX^{\mathrm{T}})=2X\\ \frac{\mathrm{d}}{\mathrm{d}X}\mathrm{tr}(BX)=B^{\mathrm{T}}=\frac{\mathrm{d}}{\mathrm{d}X}\mathrm{tr}(X^{\mathrm{T}}B^{\mathrm{T}})\\ \frac{\mathrm{d}}{\mathrm{d}X}\mathrm{tr}(X^{\mathrm{T}}AX)=(A+A^{\mathrm{T}})X\end{cases}$$

定义 6.1.4　设函数矩阵 $A(X)=(a_{lk}(X))_{m\times n}$，其中变量 $X=(x_{ij})_{p\times q}$．若每个 $a_{lk}(X)$ 对 X 皆可导，则定义函数矩阵 $A(X)$ 对 X 的微分为：

$$\mathrm{d}A(X)=\begin{pmatrix}\sum\limits_{i,j}^{p,q}\frac{\partial a_{11}(X)}{\partial x_{ij}}\mathrm{d}x_{ij} & \cdots & \sum\limits_{i,j}^{p,q}\frac{\partial a_{1n}(X)}{\partial x_{ij}}\mathrm{d}x_{ij}\\ \vdots & & \vdots\\ \sum\limits_{i,j}^{p,q}\frac{\partial a_{m1}(X)}{\partial x_{ij}}\mathrm{d}x_{ij} & \cdots & \sum\limits_{i,j}^{p,q}\frac{\partial a_{mn}(X)}{\partial x_{ij}}\mathrm{d}x_{ij}\end{pmatrix}$$

并且定义函数矩阵 $A(X)$ 对变量 X 的导数为：

$$\frac{\mathrm{d}A(X)}{\mathrm{d}X}=\begin{pmatrix}\frac{\mathrm{d}a_{11}(X)}{\mathrm{d}X} & \cdots & \frac{\mathrm{d}a_{1n}(X)}{\mathrm{d}X}\\ \vdots & & \vdots\\ \frac{\mathrm{d}a_{m1}(X)}{\mathrm{d}X} & \cdots & \frac{\mathrm{d}a_{mn}(X)}{\mathrm{d}X}\end{pmatrix}_{mp\times nq}$$

例 6.1.2　让 $X=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$，$a(X)=\begin{pmatrix}a_1(X)\\ \vdots\\ a_m(X)\end{pmatrix}$，则有

$$\frac{\mathrm{d}a^{\mathrm{T}}(X)}{\mathrm{d}X}=\begin{pmatrix}\frac{\partial a_1(X)}{\partial x_1} & \frac{\partial a_2(X)}{\partial x_1} & \cdots & \frac{\partial a_m(X)}{\partial x_1}\\ \vdots & \vdots & & \vdots\\ \frac{\partial a_1(X)}{\partial x_n} & \frac{\partial a_2(X)}{\partial x_n} & \cdots & \frac{\partial a_m(X)}{\partial x_n}\end{pmatrix}$$

$$\frac{\mathrm{d}a(X)}{\mathrm{d}X^{\mathrm{T}}}=\begin{pmatrix}\frac{\partial a_1(X)}{\partial x_1} & \frac{\partial a_1(X)}{\partial x_2} & \cdots & \frac{\partial a_1(X)}{\partial x_n}\\ \vdots & \vdots & & \vdots\\ \frac{\partial a_m(X)}{\partial x_1} & \frac{\partial a_m(X)}{\partial x_2} & \cdots & \frac{\partial a_m(X)}{\partial x_n}\end{pmatrix}$$

因此有 $\frac{\mathrm{d}a^{\mathrm{T}}(X)}{\mathrm{d}X}=\left(\frac{\mathrm{d}a(X)}{\mathrm{d}X^{\mathrm{T}}}\right)^{\mathrm{T}}$，$\frac{\mathrm{d}X}{\mathrm{d}X^{\mathrm{T}}}=I=\frac{\mathrm{d}X^{\mathrm{T}}}{\mathrm{d}X}$．

性质 6.1.5　设 A 为 $s\times m$ 常值矩阵，$f(X)$ 为向量 X 的数量函数，$a(X)$，$b(X)$ 为 X 的 m 维列向量函数，$X=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$．则

(1) $\frac{\mathrm{d}}{\mathrm{d}X}(a(X)\pm b(X))=\frac{\mathrm{d}a(X)}{\mathrm{d}X}\pm\frac{\mathrm{d}b(X)}{\mathrm{d}X}$；

(2) $\frac{\mathrm{d}}{\mathrm{d}X}(f(X)a^{\mathrm{T}}(X))=\frac{\mathrm{d}f}{\mathrm{d}X}a^{\mathrm{T}}(X)+f(X)\frac{\mathrm{d}a^{\mathrm{T}}(X)}{\mathrm{d}X}$；

(3) $\frac{\mathrm{d}}{\mathrm{d}X}(a^{\mathrm{T}}(X)b(X))=\frac{\mathrm{d}a^{\mathrm{T}}(X)}{\mathrm{d}X}b(X)+\frac{\mathrm{d}b^{\mathrm{T}}(X)}{\mathrm{d}X}a(X)$；

(4) $\frac{\mathrm{d}}{\mathrm{d}X^{\mathrm{T}}}(Aa(X))=A\cdot\frac{\mathrm{d}a(X)}{\mathrm{d}X^{\mathrm{T}}}$；

(5) $\frac{\mathrm{d}}{\mathrm{d}X^{\mathrm{T}}}(a^{\mathrm{T}}(X)b(X)) = b^{\mathrm{T}}(X)\frac{\mathrm{d}a(X)}{\mathrm{d}X^{\mathrm{T}}} + a^{\mathrm{T}}(X)\frac{\mathrm{d}b(X)}{\mathrm{d}X^{\mathrm{T}}}$.

证明: (1)、(2) 显然.现证(3). $a^{\mathrm{T}}(X)b(X) = \sum_{i=1}^{m} a_i(X)b_i(X)$,因此

$$\frac{\mathrm{d}}{\mathrm{d}X}(a^{\mathrm{T}}(X)b(X)) = \begin{pmatrix} \sum_{i=1}^{m}\left(\frac{\partial a_i(X)}{\partial x_1}b_i(X) + a_i(X)\frac{\partial b_i(X)}{\partial x_1}\right) \\ \vdots \\ \sum_{i=1}^{m}\left(\frac{\partial a_i(X)}{\partial x_n}b_i(X) + a_i(X)\frac{\partial b_i(X)}{\partial x_n}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{m}\frac{\partial a_i(X)}{\partial x_1}b_i(X) \\ \vdots \\ \sum_{i=1}^{m}\frac{\partial a_i(X)}{\partial x_n}b_i(X) \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^{m}a_i(X)\frac{\partial b_i(X)}{\partial x_1} \\ \vdots \\ \sum_{i=1}^{m}a_i(X)\frac{\partial b_i(X)}{\partial x_n} \end{pmatrix}$$

$$= \frac{\mathrm{d}a^{\mathrm{T}}(X)}{\mathrm{d}X}b(X) + \frac{\mathrm{d}b^{\mathrm{T}}(X)}{\mathrm{d}X}a(X)$$

(4)、(5) 类似可证.

□

性质 6.1.6 设 $A(X)$,$B(X)$ 是变量 $X = (x_{ij})_{p\times q}$ 的函数矩阵.若 $A(X)$,$B(X)$ 对 X 均可导,则

(1) $\frac{\mathrm{d}(A(X) \pm B(X))}{\mathrm{d}X} = \frac{\mathrm{d}A(X)}{\mathrm{d}X} \pm \frac{\mathrm{d}B(X)}{\mathrm{d}X}$;

(2) $\frac{\mathrm{d}(A(X)B(X))}{\mathrm{d}X} = \frac{\mathrm{d}A(X)}{\mathrm{d}X}(B(X)\otimes I_q) + (A(X)\otimes I_p)\frac{\mathrm{d}B(X)}{\mathrm{d}X}$.

证明: 只证(2),由定义知

$$\frac{\mathrm{d}(A(X)B(X))}{\mathrm{d}X} = \left(\frac{\mathrm{d}c_{ij}(X)}{\mathrm{d}X}\right)$$

其中 $c_{ij}(X)$ 为 $A(X)B(X)$ 的(i,j)位置元素,$c_{ij}(X) = \sum_{l} a_{il}(X)b_{lj}(X)$,注意到

$$\frac{\mathrm{d}c_{ij}(X)}{\mathrm{d}X} = \sum_{l}\left[a_{il}(X)\frac{\mathrm{d}b_{lj}(X)}{\mathrm{d}X} + \frac{\mathrm{d}a_{il}(X)}{\mathrm{d}X}b_{lj}(X)\right]$$
$$= \left(\frac{\mathrm{d}A(X)}{\mathrm{d}X}\right)_i (B(X)\otimes I_q)_j + (A(X)\otimes I_p)_i\left(\frac{\mathrm{d}B(X)}{\mathrm{d}X}\right)_j$$

式中$\left(\frac{\mathrm{d}A(X)}{\mathrm{d}X}\right)_i$ 为$\frac{\mathrm{d}A(X)}{\mathrm{d}X}$的第 i 块元行,$(B(X)\otimes I_q)_j$ 为 $B(X)\otimes I_q$ 的第 j 块元列,$(A(X)\otimes I_p)_i$ 为 $A(X)\otimes I_p$ 的第 i 块元行,$\left(\frac{\mathrm{d}B(X)}{\mathrm{d}X}\right)_j$ 为$\frac{\mathrm{d}B(X)}{\mathrm{d}X}$的第 j 块元列.因此

$$\frac{\mathrm{d}(A(X)B(X))}{\mathrm{d}X} = \frac{\mathrm{d}A(X)}{\mathrm{d}X}(B(X)\otimes I_q) + (A(X)\otimes I_p)\frac{\mathrm{d}B(X)}{\mathrm{d}X}$$

□

例 6.1.3　求 $X^{\mathrm{T}}AX$ 对 X 的导数,其中 $A\in\mathbb{R}^{n\times n}$, $X=\begin{pmatrix}x_1\\ \vdots\\ x_n\end{pmatrix}$ 为 n 维列向量变量.

解　记 $A=\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}$,则

$$\begin{aligned}\frac{\mathrm{d}(X^{\mathrm{T}}AX)}{\mathrm{d}X}&=\frac{\mathrm{d}X^{\mathrm{T}}}{\mathrm{d}X}(AX\otimes I_1)+(X^{\mathrm{T}}\otimes I_n)\frac{\mathrm{d}AX}{\mathrm{d}X}\\&=I_n(AX\otimes I_1)+(x_1I_n,\cdots,x_nI_n)\begin{pmatrix}\alpha_1^{\mathrm{T}}\\ \vdots\\ \alpha_n^{\mathrm{T}}\end{pmatrix}\\&=AX+x_1\alpha_1^{\mathrm{T}}+\cdots+x_n\alpha_n^{\mathrm{T}}\\&=AX+A^{\mathrm{T}}X\\&=(A+A^{\mathrm{T}})X\end{aligned}$$

此例实际上是给出了性质 6.1.4 最后一个等式的证明.

□

例 6.1.4　设函数矩阵 $A(X)=(a_{ij}(X))_{n\times n}$,其中 $X=(x_{lk})_{p\times q}$,且 $A(X)$ 可逆,则有

$$\frac{\mathrm{d}A^{-1}(X)}{\mathrm{d}X}=-(A^{-1}(X)\otimes I_p)\frac{\mathrm{d}A(X)}{\mathrm{d}X}(A^{-1}(X)\otimes I_q)$$

证明: 由 $A(X)A^{-1}(X)=I$ 知

$$0=\frac{\mathrm{d}(A(X)A^{-1}(X))}{\mathrm{d}X}=\frac{\mathrm{d}A(X)}{\mathrm{d}X}(A^{-1}(X)\otimes I_q)+(A(X)\otimes I_p)\frac{\mathrm{d}A^{-1}(X)}{\mathrm{d}X}$$

因此

$$\frac{\mathrm{d}A^{-1}(X)}{\mathrm{d}X}=-(A^{-1}(X)\otimes I_p)\frac{\mathrm{d}A(X)}{\mathrm{d}X}(A^{-1}(X)\otimes I_q)$$

□

以上只是把标量函数的相应概念形式上推广到函数矩阵,新内容不多.下面举两个应用的例题,通过这些例子说明引入函数矩阵的微分和积分并非是没有必要的.

例 6.1.5(欧氏空间的线性拟合)　在 m 维欧氏空间 $\mathbb{R}^m$ 中,给定 N 个点 $P_i=(x_1^{(i)},\cdots,x_m^{(i)})^{\mathrm{T}}$, $i=1,2,\cdots,N$.欲求一超平面 H: $K^{\mathrm{T}}x=k_0$,其中 $K=(k_1,\cdots,k_m)^{\mathrm{T}}$, $x=(x_1,\cdots,x_m)^{\mathrm{T}}$,使得各点 P_i 到 H 的距离平方和为最小.这样求得的 H 称为点组 $\{P_1,\cdots,P_N\}$ 的最小二乘线性拟合.特别地,当 $m=2$ 时,便是常用的平面上点组的线性拟合.

解　P_i 到 H 的距离平方为

$$\mathrm{d}_i^2=\frac{(K^{\mathrm{T}}P_i-k_0)^2}{K^{\mathrm{T}}K},i=1,2,\cdots,N$$

于是诸点 P_i 到超平面 H 的距离平方和为

$$S=\sum_{i=1}^{N}\mathrm{d}_i^2=\frac{1}{K^{\mathrm{T}}K}\sum_{i=1}^{N}(K^{\mathrm{T}}P_i-k_0)^2$$

S 可看成向量 K 和 k_0 的函数.显然 S 也是关于 $k_0,\cdots,k_m$ 的多元函数,它存在最小值.从数学分析中知,欲使 S 达到最小,一个必要条件为

$$\frac{\partial S}{\partial k_0} = -\frac{2}{K^{\mathrm{T}}K}\sum_{i=1}^{N}(K^{\mathrm{T}}P_i - k_0) = 0$$

从而有

$$k_0 = \frac{1}{N}\sum_{i=1}^{N}K^{\mathrm{T}}P_i = K^{\mathrm{T}}\left(\frac{1}{N}\sum_{i=1}^{N}P_i\right)$$

令 $\bar{P} = \frac{1}{N}\sum_{i=1}^{N}P_i$,它是 N 个点 $P_1,\cdots,P_N$ 的几何中心,从而有

$$k_0 = K^{\mathrm{T}}\bar{P}$$

$$H: K^{\mathrm{T}}(x - \bar{P}) = 0$$

这表明所要求的最佳拟合超平面一定过中心 $\bar{P}$.这时函数

$$S = \frac{1}{K^{\mathrm{T}}K}\sum_{i=1}^{N}(K^{\mathrm{T}}P_i - k_0)^2 = \frac{1}{K^{\mathrm{T}}K}\sum_{i=1}^{N}[K^{\mathrm{T}}(P_i - \bar{P})]^2$$

又令 $Y_i = P_i - \bar{P} \in \mathbb{R}^m$, $\sum_{i=1}^{N}Y_iY_i^{\mathrm{T}} = A \in \mathbb{R}^{m\times m}$,则根据 $K^{\mathrm{T}}Y_i = Y_i^{\mathrm{T}}K$ 可得

$$\begin{aligned} S &= \frac{1}{K^{\mathrm{T}}K}\sum_{i=1}^{N}(K^{\mathrm{T}}Y_i)^2 \\ &= \frac{1}{K^{\mathrm{T}}K}\sum_{i=1}^{N}K^{\mathrm{T}}Y_iY_i^{\mathrm{T}}K \\ &= \frac{1}{K^{\mathrm{T}}K}K^{\mathrm{T}}\left(\sum_{i=1}^{N}Y_iY_i^{\mathrm{T}}\right)K \\ &= \frac{1}{K^{\mathrm{T}}K}K^{\mathrm{T}}AK \end{aligned}$$

又令 $Z = \frac{K}{\|K\|}$,则 $\|Z\| = 1$,且 $S = Z^{\mathrm{T}}AZ$,于是求 $S = \frac{K^{\mathrm{T}}AK}{K^{\mathrm{T}}K}$ 的极小点 K 的问题转换为条件极值问题

$$\min_{\|Z\|=1} Z^{\mathrm{T}}AZ$$

引入 Lagrange 乘子,化为无条件极值问题,即求

$$\Phi(Z,\lambda) = Z^{\mathrm{T}}AZ + \lambda(Z^{\mathrm{T}}Z - 1)$$

的极小值点.由微分学知识,若(Z_0,λ_0)为该函数的极小值点,则有

$$\begin{cases} \left.\dfrac{\partial \Phi}{\partial Z}\right|_{(Z_0,\lambda_0)} = 2AZ_0 + 2\lambda_0 Z_0 = 0 \\ \left.\dfrac{\partial \Phi}{\partial \lambda}\right|_{(Z_0,\lambda_0)} = Z_0^{\mathrm{T}}Z_0 - 1 = 0 \end{cases}$$

于是

$$\begin{cases} AZ_0 = -\lambda_0 Z_0 \\ \|Z_0\| = 1 \\ \min S(K) = \lambda_{\min} = -\lambda_0 \end{cases}$$

其中 $\lambda_{\min}$ 为半正定阵 A 的最小特征值, Z_0 为 A 关于 $\lambda_{\min}$ 的单位特征向量. 到此为止, 我们实际上获得了要求的超平面: 它过点 $\bar{P}$, 法向量 K 为 A 的相应于 $\lambda_{\min}$ 的单位特征向量.

当半正定阵 A 的最小特征值的代数重数 $r \geqslant 1$ 时, 若对应 $\lambda_{\min}$ 线性独立的单位特征向量为 $\{Z_1, \cdots, Z_r\}$, 则过 $\bar{P}$ 以 Z_i 或 $\{Z_i\}$ 的线性组合为法方向的 H 都是最佳拟合超平面, 特别 $r > 1$ 时有无穷多个.

□

例 6.1.6(约束最小二乘问题)　给定

$$A \in \mathbb{R}^{m \times n}, \operatorname{rank} A = n. b \in \mathbb{R}^m, B \in \mathbb{R}^{k \times n}, d \in R(B)$$

求约束极小化问题 $\min\limits_{Bx=d} \| Ax - b \|^2$ 的解.

解: 显见 $\min\limits_{Bx=d} \| Ax - b \|^2$ 总是存在的. 所求问题实际上为求函数 $f(x) = \| Ax - b \|^2$ 在约束 $Bx = d$ 下的极小点和极小值. 为此, 引入 Lagrange 乘子 $2\lambda \in \mathbb{R}^k$, 化为等价的无约束极值问题

$$\varphi(x, \lambda) = \| Ax - b \|^2 + 2\lambda^{\mathrm{T}}(Bx - d)$$

若 (x_0, λ_0) 为 $\varphi(x, \lambda)$ 的极值点, 则有

$$\begin{cases} \left.\dfrac{\mathrm{d}\varphi}{\mathrm{d}x}\right|_{(x_0, \lambda_0)} = 2A^{\mathrm{T}}Ax_0 - 2A^{\mathrm{T}}b + 2B^{\mathrm{T}}\lambda_0 = 0 \\ \left.\dfrac{\mathrm{d}\varphi}{\mathrm{d}\lambda}\right|_{(x_0, \lambda_0)} = 2(Bx_0 - d) = 0 \end{cases}$$

说明极点 (x_0, λ_0) 应满足

$$\begin{pmatrix} A^{\mathrm{T}}A & B^{\mathrm{T}} \\ B & 0 \end{pmatrix} \begin{pmatrix} x_0 \\ \lambda_0 \end{pmatrix} = \begin{pmatrix} A^{\mathrm{T}}b \\ d \end{pmatrix} \tag{6.1.3}$$

由 $f(x) = \| Ax - b \|^2 = (x^{\mathrm{T}}A^{\mathrm{T}} - b^{\mathrm{T}})(Ax - b) = x^{\mathrm{T}}A^{\mathrm{T}}Ax - 2b^{\mathrm{T}}Ax + b^{\mathrm{T}}b$ 知

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{n \times n} = 2A^{\mathrm{T}}A$$

为正定阵, 因此 x_0 必为 $f(x)$ 的极值点. 在方程(6.1.3)两端左乘可逆阵

$$\begin{pmatrix} I_n & 0 \\ -B(A^{\mathrm{T}}A)^{-1} & I_k \end{pmatrix}$$

便得到

$$\begin{pmatrix} A^{\mathrm{T}}A & B^{\mathrm{T}} \\ 0 & -B(A^{\mathrm{T}}A)^{-1}B^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} x_0 \\ \lambda_0 \end{pmatrix} = \begin{pmatrix} A^{\mathrm{T}}b \\ d - B(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b \end{pmatrix}$$

结果得

$$\begin{aligned} \lambda_0 &= (B(A^{\mathrm{T}}A)^{-1}B^{\mathrm{T}})^{\{1\}}(B(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b - d) \\ &\quad + (I_k - (B(A^{\mathrm{T}}A)^{-1}B^{\mathrm{T}})^{\{1\}}(B(A^{\mathrm{T}}A)^{-1}B^{\mathrm{T}} \end{aligned}$$

其中 y 为 IR^k 中的任意元素.

$$x_0 = (A^{\mathrm{T}}A)^{-1}(A^{\mathrm{T}}b - B^{\mathrm{T}}\lambda_0)$$

$$f(x_0) = \min_{Bx=d} \| Ax - b \|^2 = \| Ax_0 - b \|^2 = \| A(A^{\mathrm{T}}A)^{-1}(A^{\mathrm{T}}b - B^{\mathrm{T}}\lambda_0) - b \|^2$$

□

6.2 函数矩阵微积分的应用

前一节,我们建立了函数矩阵微积分的基本理论.本节进一步介绍函数矩阵微积分的几个应用.

(1)函数向量的线性相关性判定

设 $\alpha_i(x)=(\alpha_{1i}(x),\cdots,\alpha_{ni}(x))^{\mathrm{T}}, i=1,2,\cdots,m$,为定义在区域 D 上的函数向量.若存在不全为 0 的数 $c_1,\cdots,c_m$ 使得

$$c_1\alpha_1(x)+\cdots+c_m\alpha_m(x)=0, \forall x\in D$$

则称 $\alpha_1(x),\cdots,\alpha_m(x)$在 D 上**强线性相关**;当 $\alpha_1(x),\cdots,\alpha_m(x)$不是强线性相关时,称为**弱线性无关**.一般地,我们有:

定义 6.2.1 设 $\alpha_1(x),\cdots,\alpha_m(x)$是一组同维的函数向量,若在区域 D 上存在不全为 0 的函数 $\varphi_1(x),\cdots,\varphi_m(x)$使得

$$\varphi_1\alpha_1(x)+\cdots+\varphi_m\alpha_m(x)\equiv 0, \forall x\in D$$

则称 $\alpha_1(x),\cdots,\alpha_m(x)$在区域 D 上**线性相关**;否则称为在区域 D 上**线性无关**.

定义 6.2.2 设 $A(x)$是在区域 D 上的 $m\times n$ 函数阵,称其不恒为 0 的子式的最高阶为 $A(x)$**的秩**,记为 $\mathrm{rank}A(x)$或秩 $A(x)$.

易知,利用定义 6.2.1,我们还可定义函数向量组的秩等概念.

定义 6.2.3 设 $\alpha_1(x),\cdots,\alpha_m(x)$是区间$[a,b](a,b\in\mathbb{R})$上连续的 n 维列函数向量,记

$$g_{ij}=\int_a^b\alpha_i^{\mathrm{T}}(x)\alpha_j(x)\mathrm{d}x, \quad i,j=1,2,\cdots,m \tag{6.2.1}$$

称 $G=(g_{ij})_{m\times m}$为**该函数向量的 Gram 矩阵**.

定理 6.2.1 在$[a,b]$上连续的同维列函数向量 $\alpha_1(x),\cdots,\alpha_m(x)$弱线性无关的充要条件是它的 Gram 矩阵可逆.

证明:设 $k_1\alpha_1(x)+\cdots+k_m\alpha_m(x)=0$,其中 $k_1,\cdots,k_m$ 为常数,于是

$$k_1\int_a^b\alpha_i^{\mathrm{T}}(x)\alpha_1(x)\mathrm{d}x+\cdots+k_m\int_a^b\alpha_i^{\mathrm{T}}(x)\alpha_m(x)\mathrm{d}x=0, \quad i=1,2,\cdots,m$$

即

$$\sum_{j=1}^m k_jg_{ij}=0, \quad i=1,2,\cdots,m$$

令 $k=\begin{pmatrix}k_1\\ \vdots\\ k_m\end{pmatrix}$,得

$$GK=0$$

$\Leftarrow$　当 G 可逆时,$GX=0$ 关于 X 只有零解,因此 $\alpha_1(x),\cdots,\alpha_m(x)$不可能强线性相关.

$\Rightarrow$　反证,若 G 不可逆,则 $GX=0$ 关于 X 有非零解 $X_0=(k_1,\cdots,k_m)^{\mathrm{T}}$,令

$$\alpha(x)=k_1\alpha_1(x)+\cdots+k_m\alpha_m(x)$$

则有

$$\begin{aligned}\int_a^b\alpha^{\mathrm{T}}(x)\alpha(x)\mathrm{d}x &= \int_a^b\sum_{i,j}k_ik_j\alpha_i^{\mathrm{T}}(x)\alpha_j(x)\mathrm{d}x\\ &= \sum_{i,j}k_ik_j\int_a^b\alpha_i^{\mathrm{T}}(x)\alpha_j(x)\mathrm{d}x\\ &= X_0^{\mathrm{T}}GX_0 = 0\end{aligned}$$

于是 $\alpha(x)=0$,此时 $\alpha_1(x),\cdots,\alpha_m(x)$强线性相关.

□

定义 6.2.4　设 $\alpha_i(x)=(\alpha_{i1}(x),\cdots,\alpha_{in}(x))^{\mathrm{T}},i=1,2,\cdots,m$,是区间$[a,b]$上有 $m-1$ 阶导数的函数向量.记

$$A(x)=\begin{pmatrix}\alpha_1^{\mathrm{T}}(x)\\ \vdots\\ \alpha_m^{\mathrm{T}}(x)\end{pmatrix}=(\alpha_{ij}(x))_{m\times n}$$

令

$$W(x)=(A(x),A(x),\cdots,A^{(m-1)}(x))_{m\times mn}$$

称 $W(x)$是 $\alpha_1(x),\cdots,\alpha_m(x)$的 **Wronski 矩阵(朗斯基矩阵)**.

定理 6.2.2　设 $W(x)$是列向量 $\alpha_1(x),\cdots,\alpha_m(x)$的 Wronski 矩阵 . 若存在 $x_0\in[a,b]$使得 $\operatorname{rank}W(x_0)=m$,则向量 $\alpha_1(x),\cdots,\alpha_m(x)$在$[a,b]$上弱线性无关.

证明: 反证.设 $\alpha_1(x),\cdots,\alpha_m(x)$在$[a,b]$上强线性相关,则存在 m 维的非零常数列 C 使得

$$(\alpha_1(x),\cdots,\alpha_m(x))C=0$$

稍作变形后有

$$C^{\mathrm{T}}\begin{pmatrix}\alpha_1^{\mathrm{T}}(x)\\ \vdots\\ \alpha_m^{\mathrm{T}}(x)\end{pmatrix}=0$$

亦即

$$C^{\mathrm{T}}A(x)\equiv 0,\quad \forall x\in[a,b]$$

由此得

$$\begin{gathered}C^{\mathrm{T}}A'(x)\equiv 0,\quad \forall x\in[a,b]\\ \vdots\\ C^{\mathrm{T}}A^{(m-1)}(x)\equiv 0,\quad \forall x\in[a,b]\end{gathered}$$

故有

$$C^{\mathrm{T}}W(x_0)=C^{\mathrm{T}}(A(x_0),\cdots,A^{(m-1)}(x_0))=0$$

注意到 $W(x_0)$为行满秩阵,右可消,因此 $C=0$,矛盾.

□

(2)线性时变系统的状态方程

线性时变系统的状态方程,即为变系数的线性微分方程组,其一般形式是

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)+B(t)u(t) \tag{6.2.2}$$

其中 $A(t),B(t)$分别为已知的 $n\times n,n\times m$ 的函数矩阵,且 $A(t)$与 $B(t)$均在所讨论区域 Ω 上连续;$x(t)$为 $n\times m$ 的未知函数矩阵,而 $u(t)$为 $m\times m$ 的已知函数矩阵.当方程(6.2.2)中常项 $B(t)u(t)$为 0 时,称之为**变系数的齐次微分方程组**.相应地,常项 $B(t)u(t)$不为 0 时,则称之为**变系数的非齐次微分方程组或非齐次时变系统状态方程**.

以下将利用函数矩阵的相关性质分别给出这两类时变系统状态方程定解问题的具体求解.首先考虑齐次的情形:

$$\begin{cases}\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)\\ x(t)\big|_{t=t_0}=x(t_0)\end{cases} \tag{6.2.3}$$

由定理 6.1.1 知,该定解问题存在唯一解.为找出该解,我们将之转换为求时变系统$\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)$的所谓转移矩阵.转移矩阵的定义如下:

定义 6.2.5 设 n 个关于 t 的函数向量

$$\phi_j(t,t_0)=\begin{pmatrix}x_{1j}(t,t_0)\\ \vdots\\ x_{nj}(t,t_0)\end{pmatrix},\quad j=1,2,\cdots,n$$

满足条件

$$\begin{cases}\dfrac{\mathrm{d}\phi_j}{\mathrm{d}t}=A(t)\phi_j\\ j=1,2,\cdots,n\end{cases}\quad 且\ \phi_j(t,t_0)\big|_{t=t_0}=\begin{pmatrix}x_{1j}(t_0,t_0)\\ \vdots\\ x_{nj}(t_0,t_0)\end{pmatrix}=\begin{pmatrix}\vdots\\ 1\\ 0\\ \vdots\end{pmatrix}=e_j,j=1,2,\cdots,n$$

则称 n 阶方阵

$$\phi(t,t_0)=(\phi_1,\cdots,\phi_n)=\begin{pmatrix}x_{11}(t,t_0)&\cdots&x_{1n}(t,t_0)\\ \vdots&&\vdots\\ x_{n1}(t,t_0)&\cdots&x_{nn}(t,t_0)\end{pmatrix}$$

为方程组$\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)$的**转移矩阵或基本矩阵**.

有时,对一个 n 阶函数方阵,只要其每个列向量都是$\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)$的解,也称它为$\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)$的转移矩阵.但我们指出,本书所涉及的转移矩阵统一指为定义 6.2.5 所规定的情形.

例 6.2.1　令 $A(t)$，$\phi_1(t,t_0)$及 $\phi_2(t,t_0)$分别为

$$A(t)=\begin{pmatrix}2 & -e^t\\ e^{-t} & 1\end{pmatrix},\phi_1(t,t_0)=\begin{pmatrix}e^{2(t-t_0)}\cos(t-t_0)\\ e^{t-t_0}\sin(t-t_0)\end{pmatrix}$$

$$\phi_2(t,t_0)=\begin{pmatrix}-e^{2t-t_0}\sin(t-t_0)\\ -e^{t-t_0}\cos(t-t_0)\end{pmatrix}$$

则有

$$\frac{d\phi_i(t,t_0)}{dt}=A(t)\phi_i(t,t_0),i=1,2$$

$$\phi_i(t_0,t_0)=e_i,i=1,2$$

故矩阵 $\phi(t,t_0)=(\phi_1,\phi_2)$为$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵.

□

给定时变系统$\frac{dx(t)}{dt}=A(t)x(t)$，现在一个基本的问题是，其转移矩阵是否存在？如果存在的话，又是否唯一？转移矩阵的下述性质 6.2.1 给出了它们肯定的答案.

性质 6.2.1　n 阶方阵$\phi(t,t_0)$为时变系统$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵的充要条件是 $\phi(t,t_0)$为$\begin{cases}\frac{dx(t)}{dt}=A(t)x(t)\\ x(t)|_{t=t_0}=I_n\end{cases}$的解.

因此，利用定理 6.1.1 得知，转移矩阵总是存在且唯一的. 为建立方程(6.2.3)的解与转移矩阵之间的关系，我们有：

性质 6.2.2　设 $\phi(t,t_0)$为$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵，则方程(6.2.3)的解为$x(t)=\phi(t,t_0)x(t_0)$.

证明： 直接验证即可.

□

此性质还表明，若 $x(t_0)$代表系统 t_0 时刻的状态(即为初始状态)，$x(t)$代表系统在 t 时刻的状态，则对初始状态左乘 $\phi(t,t_0)$便得到系统在 t 时刻的状态. 这正说明了将 $\phi(t,t_0)$称为转移矩阵的原因，因此也说明了求转移矩阵的重要性. 换句话说，方程(6.2.3)的求解关键就在于求转移矩阵 $\phi(t,t_0)$.

除了性质 6.2.1 及性质 6.2.2 外，转移矩阵还有下列性质.

性质 6.2.3　设 $\phi(t,t_0)$为时变系统$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵，则

(1) $\phi(t,t_0)=\phi(t,t_1)\phi(t_1,t_0)$；

(2) $\phi^{-1}(t,t_0)=\phi(t_0,t)$；

(3) $\det\phi(t,t_0)=e^{\int_{t_0}^{t}\operatorname{tr}A(\tau)d\tau}$(Jacobi 恒等式).

证明： (1)由性质 6.2.2 知，定解问题(6.2.3)的唯一解为

$$x(t)=\phi(t,t_0)x(t_0)$$

由此,系统在 t_1 时刻的状态为 $x(t_1)=\phi(t_1,t_0)x(t_0)$,注意到 $x(t)=\phi(t,t_1)x(t_1)$是定解问题

$$\begin{cases}\dfrac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)\\ x(t)\big|_{t=t_1}=x(t_1)\end{cases} \tag{6.2.4}$$

的唯一解.因此有 $x(t)=\phi(t,t_0)x(t_0)$与 $x(t)=\phi(t,t_1)x(t_1)$均是方程(6.2.4)的解.结果

$$\phi(t,t_0)x(t_0)=\phi(t,t_1)\phi(t_1,t_0)x(t_0)$$

由 $x(t_0)$的任意性知:$\phi(t,t_0)=\phi(t,t_1)\phi(t_1,t_0)$.获证.

(2)由 $\phi(t,t_0)=\phi(t,t_1)\phi(t_1,t_0)$,两边让 $t=t_0$,得、

$$I=\phi(t_0,t_1)\phi(t_1,t_0)$$

亦即

$$\phi^{-1}(t,t_0)=\phi(t_0,t)$$

(3)将 $\phi(t,t_0)$进行行分块,$\phi(t,t_0)=\begin{pmatrix}X_1(t)\\ \vdots\\ X_n(t)\end{pmatrix}$,据$\dfrac{\mathrm{d}\phi(t,t_0)}{\mathrm{d}t}=A(t)\phi(t,t_0)$得:

$$\frac{\mathrm{d}X_j}{\mathrm{d}t}=[A(t)\phi(t,t_0)]_j=\sum_{l=1}^{n}a_{jl}(t)X_l(t),j=1,2,\cdots,n$$

根据行列式的求导法则有:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}[\det\phi(t,t_0)]&=\sum_{i=1}^{n}\det\begin{pmatrix}X_1(t)\\ \vdots\\ X_{i-1}(t)\\ \dfrac{\mathrm{d}X_i(t)}{\mathrm{d}t}\\ \vdots\\ X_n(t)\end{pmatrix}=\sum_{i=1}^{n}\det\begin{pmatrix}X_1(t)\\ \vdots\\ X_{i-1}(t)\\ \sum\limits_{l=1}^{n}a_{il}(t)X_l(t)\\ \vdots\\ X_n(t)\end{pmatrix}\\ &=\sum_{i=1}^{n}a_{ii}(t)\det\phi(t,t_0)\\ &=\operatorname{tr}A(t)\cdot\det\phi(t,t_0)\end{aligned}$$

从而得

$$\varphi=\det\phi(t,t_0)=e^{\int_{t_0}^{t}\operatorname{tr}A(\tau)\mathrm{d}\tau}$$

□

接下来,重点关注转移矩阵的求法.

方法一 (迭代格式法) 该方法的本质是定理6.1.1的证明中出现的下列迭代格式:

$$\begin{cases} X_0 = I_n \\ X_{k+1} = I_n + \int_{t_0}^{t} A(\tau)X_k(\tau)\mathrm{d}\tau \\ \phi(t,t_0) = \lim\limits_{k} X_k(t) \end{cases}$$

该方法的特点是:尽管取得的是一个近似解,但其精度容易掌握.

方法二　(Peano - Baker 级数法)　一般地,设 $A(t)$ 为在区域 Ω 上连续的 n 阶函数方阵,称级数

$$I + \int_{t_0}^{t} A(v_1)\mathrm{d}v_1 + \int_{t_0}^{t} A(v_1)\mathrm{d}v_1\int_{t_0}^{v_1} A(v_2)\mathrm{d}v_2 + \int_{t_0}^{t} A(v_1)\mathrm{d}v_1\int_{t_0}^{v_1} A(v_2)\mathrm{d}v_2\int_{t_0}^{v_2} A(v_3)\mathrm{d}v_3 + \cdots$$

为 Peano - Baker 级数. Peano - Baker 级数法求转移矩阵的主要根据是如下定理:

定理 6.2.3　设时变系统为 $\dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = A(t)x(t)$,则 Peano - Baker 级数的和矩阵便为该时变系统的转移矩阵.

证明:(1) 首先证明在定理的假设条件下,Peano - Baker 级数在 $[t_0,t_1]$ 上一致收敛,为此记 $A(t) = (a_{ij}(t))_{n\times n}$,$M$ 为固定的整数使得

$$|a_{ij}(t)| \leqslant M,\quad \forall t \in [t_0,t_1], i,j = 1,2,\cdots,n$$

又记 $a_k^{(i,j)} = \int_{t_0}^{t} a_{ij}(v_1)\mathrm{d}v_1\int_{t_0}^{v_1} a_{ij}(v_2)\mathrm{d}v_2\cdots\int_{t_0}^{v_{k-1}} a_{ij}(v_k)\mathrm{d}v_k$, $k = 1,2,\cdots$,于是有

$$|a_1^{(i,j)}| = \left|\int_{t_0}^{t} a_{ij}(v_1)\mathrm{d}v_1\right| \leqslant M(t_1 - t_0) = \frac{1}{1!}M(t_1 - t_0)$$

$$|a_2^{(i,j)}| \leqslant \int_{t_0}^{t} |a_{ij}(v_1)| \cdot \int_{t_0}^{v_1} a_{ij}(v_2)\mathrm{d}v_2 |\mathrm{d}v_1 \leqslant \int_{t_0}^{t} M^2(v_1 - t_0)\mathrm{d}v_1 = \frac{1}{2!}M^2(t - t_0)^2$$

类似地有:

$$|a_k^{(i,j)}| \leqslant \int_{t_0}^{t} \frac{1}{(k-1)!}M^k(v_1 - t_0)^{k-1}\mathrm{d}v_1 = \frac{1}{k!}M^k(t - t_0)^k$$

$$\vdots$$

因此,Peano - Baker 级数一致收敛.

(2) 现设 Peano - Baker 级数一致收敛于 $\phi(t,t_0)$,于是

$$\begin{aligned} \frac{\mathrm{d}\phi(t,t_0)}{\mathrm{d}t} &= A(t) + A(t)\int_{t_0}^{t} A(v_2)\mathrm{d}v_2 + A(t)\int_{t_0}^{t} A(v_2)\mathrm{d}v_2\int_{t_0}^{v_2} A(v_3)\mathrm{d}v_3 + \cdots \\ &= A(t)[I + \int_{t_0}^{t} A(v_2)\mathrm{d}v_2 + \int_{t_0}^{t} A(v_2)\mathrm{d}v_2\int_{t_0}^{v_2} A(v_3)\mathrm{d}v_3 + \cdots] \\ &= A(t) \cdot \phi(t,t_0) \end{aligned}$$

和

$$\phi(t_0,t_0) = I_n$$

所以 $\phi(t,t_0)$ 为时变系统 $\dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = A(t)x(t)$ 的转移矩阵.

□

在实际应用中,由于时变系统$\frac{dx(t)}{dt}=A(t)x(t)$中的$A(t)$还会具有某些特殊的性质,因而此时转移矩阵的获得也往往还可找到更为便捷的途径,见下例:

例 6.2.2 设$A(t)$为在区域Ω上连续的n阶函数方阵,且

$$\int_{t_0}^{t}A(v)\mathrm{d}v\cdot A(t)=A(t)\cdot\int_{t_0}^{t}A(v)\mathrm{d}v$$

则$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵为$e^{\int_{t_0}^{t}A(v)dv}$,特别地,若$\forall t_1,t_2\in\Omega$均有$A(t_1)A(t_2)=A(t_2)A(t_1)$,则$e^{\int_{t_0}^{t}A(v)dv}$即为$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵.

证明:事实上,由$e^{\int_{t_0}^{t}A(v)dv}$的幂级数展式

$$e^{\int_{t_0}^{t}A(v)dv}=I+\int_{t_0}^{t}A(v)\mathrm{d}v+\frac{1}{2!}\left(\int_{t_0}^{t}A(v)\mathrm{d}v\right)^2+\cdots$$

两边对t求导得:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}e^{\int_{t_0}^{t}A(v)dv}&=A(t)\left[I+\int_{t_0}^{t}A(v)\mathrm{d}v+\frac{1}{2!}\left(\int_{t_0}^{t}A(v)\mathrm{d}v\right)^2+\cdots\right]\\&=A(t)\cdot e^{\int_{t_0}^{t}A(v)dv}\end{aligned}$$

并注意到

$$e^{\int_{t_0}^{t_0}A(v)dv}=I$$

可知$e^{\int_{t_0}^{t}A(v)dv}$为$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵.

□

前面所做使我们对齐次时变系统的定解问题有了较好的回答.和标量的常微分方程一样,采用常数变易法,能轻松地解决非齐次时变系统的定解问题:

$$\begin{cases}\frac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)x(t)+B(t)u(t)\\x(t)|_{t=t_0}=x(t_0)\end{cases}\tag{6.2.5}$$

式中$A(t),B(t),u(t),x(t),x(t_0)$如前,未知量为$x(t)$,已知矩阵$A(t),B(t)$均在所求讨论区域$\Omega$上连续.

定理 6.2.4 方程(6.2.5)有唯一解,该唯一解为

$$x(t)=\phi(t,t_0)\left[x(t_0)+\int_{t_0}^{t}\phi(t_0,v)B(v)u(v)\mathrm{d}v\right]$$

其中$\phi(t,t_0)$为齐次时变系统$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵.

证明:首先我们用常数变易法求出方程(6.2.5)的一个解.设$x(t)=\phi(t,t_0)C(t)$为方程(6.2.5)的一个解,其中$\phi(t,t_0)$为$\frac{dx(t)}{dt}=A(t)x(t)$的转移矩阵,$C(t)$为待定的函数向量.于是

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \frac{\mathrm{d}\phi(t,t_0)}{\mathrm{d}t}C(t) + \phi(t,t_0)\frac{\mathrm{d}C(t)}{\mathrm{d}t}$$
$$= A(t)\phi(t,t_0)C(t) + \phi(t,t_0)C'(t)$$

代入方程(6.2.5)得

$$A(t)\phi(t,t_0)C(t) + B(t)u(t) = A(t)\phi(t,t_0)C(t) + \phi(t,t_0)C'(t)$$

即有

$$C'(t) = \phi^{-1}(t,t_0)B(t)u(t) = \phi(t_0,t)B(t)u(t)$$

因此

$$C(t) = C + \int_{t_0}^{t}\phi(t_0,v)B(v)u(v)\mathrm{d}v$$

由初始条件$x(t)|_{t=t_0} = x(t_0)$有

$$x(t_0) = \phi(t_0,t_0)C(t_0) = C$$

故方程(6.2.5)的一个解为

$$x(t) = \phi(t,t_0)[x(t_0) + \int_{t_0}^{t}\phi(t_0,v)B(v)u(v)\mathrm{d}v]$$

最后,若 $x(t),y(t)$ 均为方程(6.2.5)的解,则 $z(t) = y(t) - x(t)$ 便为方程

$$\begin{cases}\dfrac{\mathrm{d}z(t)}{\mathrm{d}t} = A(t)z(t)\\ z(t)|_{t=t_0} = 0\end{cases} \tag{6.2.6}$$

的解.由定理 6.1.1 及性质 6.2.2 知,方程(6.2.6)的解只有一个,它为

$$\phi(t,t_0)z(t_0) = 0$$

因此,$x(t) = y(t),\forall t \in \Omega$.

□

例 6.2.3　对非齐次定常时变系统

$$\begin{cases}\dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = Ax(t) + B(t)u(t)\\ x(t)|_{t=t_0} = x(t_0)\end{cases}$$

由于 A 为常值 n 阶方阵,因此,由例 6.2.2 知其转移矩阵即为

$$\phi(t,t_0) = \mathrm{e}^{A(t-t_0)}$$

此时,该非齐次定常时变系统的解(只有 1 个)为

$$x(t) = \mathrm{e}^{A(t-t_0)}x(t_0) + \int_{t_0}^{t}\mathrm{e}^{A(t-v)}B(v)u(v)\mathrm{d}v$$

□

例 6.2.4　求下列定解问题的解:

$$\begin{cases}\dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = A(t)x(t) + B(t)u(t)\\ x(t)|_{t=t_0} = x_0\end{cases}$$

其中 $A(t)=\begin{pmatrix}1 & e^{-t}\\ 0 & 1\end{pmatrix}$，$B(t)u(t)=\begin{pmatrix}t\\1\end{pmatrix}$，$x_0=\begin{pmatrix}1\\-1\end{pmatrix}$，$t_0=0$.

解：由 $A(t_1)A(t_2)=A(t_2)A(t_1)$ 对 $\forall t_1,t_2$ 均成立知，转移矩阵即为

$$\phi(t,0)=\phi(t,t_0)=e^{\int_{t_0}^{t}A(v)dv}=e^{\begin{pmatrix}t-t_0 & e^{-t_0}-e^{-t}\\ 0 & t-t_0\end{pmatrix}}=e^{\begin{pmatrix}t & 1-e^{-t}\\ 0 & t\end{pmatrix}}$$

由矩阵函数的计算可得

$$\phi(t,0)=\begin{pmatrix}e^{t} & e^{t}-1\\ 0 & e^{t}\end{pmatrix}$$

另一方面，计算 $\int_{t_0}^{t}\phi(t_0,v)B(v)u(v)dv$ 有

$$\int_0^t\begin{pmatrix}e^{t-v} & e^{t-2v}-e^{-v}\\ 0 & e^{t-v}\end{pmatrix}\begin{pmatrix}v\\1\end{pmatrix}dv=\int_0^t\begin{pmatrix}ve^{t-v}-e^{-v}+e^{t-2v}\\ e^{t-v}\end{pmatrix}dv$$

$$=\begin{pmatrix}\int_0^t(ve^{t-v}-e^{-v}+e^{t-2v})dv\\ \int_0^t e^{t-v}dv\end{pmatrix}=\begin{pmatrix}1-te^{-t}-\frac{1}{2}e^{-2t}\\ 1-e^{-t}\end{pmatrix}$$

最后定解问题的解为

$$x(t)=\phi(t,0)\left[x(0)+\begin{pmatrix}1-te^{-t}-\frac{1}{2}e^{-2t}\\ 1-e^{-t}\end{pmatrix}\right]=\begin{pmatrix}2e^{t}+\frac{1}{2}e^{-t}-t-1\\ -1\end{pmatrix}$$

□

6.3 稳定性与李雅普诺夫定理

对齐次时变系统状态方程

$$\begin{cases}\dfrac{dX}{dt}=A(t)X(t)\\ X(t_0)=C\end{cases}\tag{6.3.1}$$

记该系统的转移矩阵为 $\phi(t,t_0)$，则该方程的唯一解为 $X(t)=\phi(t,t_0)C$，它既是 t 的函数，又依赖于初始状态 C.现假设初始状态 C 进行稍微的扰动，那么它会对解 $X(t)$ 产生怎样的影响呢？本节主要讨论此问题.为凸显初始状态 $C\in\mathbb{C}^{n\times m}$ 对解 $X(t)$ 的影响，特记 $X_C(t)=\phi(t,t_0)C$，且把 $X_0(t)=\phi(t,t_0)\cdot 0=0$ 称作**方程的零解**.给定 $\mathbb{C}^{n\times m}$ 的一个相容矩阵范数 $\|\cdot\|$，为讨论初始状态 C 的扰动对解 $X(t)$ 产生的影响，引入如下两个定义：

定义 6.3.1 称方程(6.3.1)的零解在李雅普洛夫意义下**稳定**，是指：

$\forall\varepsilon>0$，$\exists\delta(\varepsilon)>0$，使得当 $\|C\|<\delta(\varepsilon)$ 时，对一切 $t>t_0$ 均有 $\|X_C(t)\|\leqslant\varepsilon$.

定义 6.3.2　称方程(6.3.1)的零解在李雅普洛夫意义下是**渐近稳定**的是指:方程(6.3.1)的零解在李雅普洛夫意义下稳定且满足$\lim\limits_{t\to\infty}X_C(t)=0$, $\forall\ C\in\mathbb{C}^{n\times m}$.

容易看出,上述定义与$\mathbb{C}^{n\times m}$中相容矩阵范数$\|\cdot\|$的选取无关.因此在证明与此相关的论断过程中,常根据需要,假定相容矩阵范数$\|\cdot\|$为 F 范数或谱范数等一些我们熟悉的范数.

为解决方程(6.3.1)零解在李雅洛夫意义下的稳定性与渐近稳定性问题,我们有下述定理:

定理 6.3.1　(1)方程(6.3.1)的零解在李雅洛夫意义下是稳定的当且仅当对一切 $t\geqslant t_0$,存在常数 $r>0$,使得$\|\phi(t,t_0)\|_F\leqslant r$,即 $A(t)$的转移矩阵是一致有界的.

(2)方程(6.3.1)的零解在李雅洛夫意义下是渐近稳定的当且仅当$\lim\limits_{t\to\infty}\phi(t,t_0)=0$.

证明:(1)$\Rightarrow$　反证,$\phi(t,t_0)$不是一致有界的,于是 $\phi(t,t_0)$中存在元素 $\phi_{ij}(t)$在$[t_0,+\infty)$上无界,此时取初始状态 $C=(ae_j,0,\cdots,0)_{n\times m}$, $a\neq 0$, $X_C(t)=\phi(t,t_0)C$ 中$(i,1)$的位置元素为 $a\phi_{ij}(t)$,明显地,$a\phi_{ij}(t)$在$[t_0,+\infty)$仍无界.矛盾于零解的稳定性.

$\Leftarrow$　注意到 $X_C(t)=\phi(t,t_0)C$,因此$\forall\ \varepsilon>0$,取 $\delta=\dfrac{\varepsilon}{r}$,则有当$\|C\|_F<\delta$ 时,对一切 $t\geqslant t_0$ 均有$\|X_C(t)\|_F\leqslant\|\phi(t,t_0)\|_F\|C\|_F\leqslant\delta\cdot r=\varepsilon$.

(2)$\Rightarrow$　在 $X_0(t)$是李雅普洛夫意义上渐近稳定的情况下,若$\lim\limits_{t\to\infty}\phi(t,t_0)\neq 0$,则 $\phi(t,t_0)$中至少存在一个元素 $\phi_{ij}(t)$,当 $t\to+\infty$时,它不收敛于 0.取 $C=(e_j,0,\cdots,0)$,此时 $\phi(t,t_0)C$ 中$(i,1)$位置元素为 $\phi_{ij}(t)$,它不收敛于 0(当 $t\to+\infty$时),矛盾于$\lim\limits_{t\to\infty}X_C(t)=0$.

$\Leftarrow$　是显然的.

□

例 6.3.1　$\begin{cases}\dfrac{dX}{dt}=X\\ X(t_0)=C\end{cases}$ 的零解是不稳定的;$\begin{cases}\dfrac{dX}{dt}=X(t)\sin t\\ X(t_0)=C\end{cases}$ 的零解是稳定的,但不是渐近稳定的;$\begin{cases}\dfrac{dX}{dt}=-X\\ X(t_0)=C\end{cases}$ 的零解是渐近稳定的.

当方程(6.3.1)中的 $A(t)$为常数矩阵 A 时,李雅普洛夫将其零解渐近稳定性的判定与 A 的特征值刻画联系起来,给出了著名的李雅普洛夫判据:

定理 6.3.2　当 A 是常数矩阵时,方程$\dfrac{dX}{dt}=AX(t)$的零解是渐近稳定的当且仅当:A 的所有特征值有负实部;方程$\dfrac{dX}{dt}=AX(t)$的零解是稳定的当且仅当 A 的所有特征值均有非正的实部并且实部为零的特征值所对应的 Jordan 块是一阶的.

证明:当 A 为常数矩阵时,A 的转移矩阵$\phi(t,t_0)=e^{A(t-t_0)}$, $X_C(t)=e^{A(t-t_0)}C$.取 A

的 Jordan 分解 $A=P\begin{pmatrix}J_1&&\\&\ddots&\\&&J_t\end{pmatrix}P^{-1}$，于是 $e^{A(t-t_0)}=P\begin{pmatrix}B_1&&\\&\ddots&\\&&B_t\end{pmatrix}P^{-1}$，其中

$$B_i=e^{\lambda_i(t-t_0)}\begin{pmatrix}1 & t-t_0\cdots & \dfrac{(t-t_0)^{n_i-1}}{(n_i-1)!}\\ & \ddots\ \ddots & \vdots\\ & & t-t_0\\ & \ddots & \\ & & 1\end{pmatrix}$$

因此 $\lim\limits_{t\to+\infty}e^{A(t-t_0)}=0\Leftrightarrow$对每一块 B_i，均有 $\lim\limits_{t\to+\infty}B_i=0\Leftrightarrow$每个 λ_i 的实部为负数.

$e^{A(t-t_0)}$一致有界$\Leftrightarrow$每个 B_i 均是一致有界$\Leftrightarrow$对每个 λ_i 而言，要么 λ_i 的实部为负数；要么 λ_i 的实部为 0，但此时对所对应的 Jordan 块是一阶的.

□

一般地，对一个复方阵 A，若它的所有特征值均有负实部，我们称 A 为**稳定矩阵**；若它的所有特征值均有非正的实部且实部为零的特征值所对应的 Jordan 块是一阶的，则我们称 A 是一个**半稳定矩阵**.

李雅普洛夫判据表明：常系数时变系统状态方程$\dfrac{dX(t)}{dt}=AX(t)$，其零解的稳定性与渐近稳定性分别对应于 A 是半稳定矩阵与稳定矩阵.接下来，我们讨论常数矩阵 A 的稳定性判定问题.对实方阵 A，李雅普洛夫在此问题上也给出了深刻的结果：

定理 6.3.3 设 $A\in\mathbb{R}^{n\times n}$，则以下两条等价：

(1) A 是稳定矩阵；

(2)关于 V 的矩阵方程 $A^TV+VA=-I$ 有唯一的实正定对称解.

证明：(1)⇒(2) 构造$\mathbb{R}$－线性映射如下：

$$F:\mathbb{R}^{n\times n}\to\mathbb{R}^{n\times n}$$
$$M\to A^TM+MA$$

验证 $N(F)=\{0\}$.任取 $Z\in N(F)$，则

$$A^TZ+ZA=0$$
$$-A^TZ=ZA$$
$$(-tA^T)^kZ=Z(tA)^k,\ \forall k\in\mathbb{N},\forall t\in\mathbb{R}$$
$$e^{-tA^T}Z=Ze^{tA}$$
$$Z=e^{tA^T}Ze^{tA}$$
$$Z=\lim_{t\to+\infty}Z=\lim_{t\to+\infty}e^{tA^T}Ze^{tA}=0$$

因此，F 是可逆映射.特别地，$F^{-1}(-I)$有唯一解 V，即$\exists!$ V使得 $A^TV+VA=-I$.注意到，V^T 也满足矩阵方程 $A^TX+XA=-I$，因此 $V=V^T$，V 为对称矩阵.

下证 V 为正定矩阵.反设 V 不为正定阵,则可取 $c\in\mathbb{R}^n$ 且 $c\neq 0$,使得 $c^{\mathrm{T}}Vc\leqslant 0$,于是让 $x=\mathrm{e}^{At}\cdot c$,并构造二次型 $f(t)=x^{\mathrm{T}}Vx=c^{\mathrm{T}}\cdot(\mathrm{e}^{At})^{\mathrm{T}}\cdot V\cdot\mathrm{e}^{At}\cdot c$,直接计算知

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}t}&=\frac{\mathrm{d}x^{\mathrm{T}}Vx}{\mathrm{d}t}=\frac{\mathrm{d}x^{\mathrm{T}}}{\mathrm{d}t}Vx+x^{\mathrm{T}}V\cdot\frac{\mathrm{d}x}{\mathrm{d}t}\\&=c^{\mathrm{T}}\mathrm{e}^{A^{\mathrm{T}}t}A^{\mathrm{T}}Vx+x^{\mathrm{T}}V\cdot A\mathrm{e}^{At}\cdot c\\&=x^{\mathrm{T}}(A^{\mathrm{T}}V+VA)x\\&=x^{\mathrm{T}}(-I)x=-x^{\mathrm{T}}x<0\end{aligned}$$

这表明 $f(t)$ 对一切 $t\in\mathbb{R}$ 都是严格递减的,特别地有

$$f(1)<f(0)=c^{\mathrm{T}}Vc\leqslant 0$$

而对一切 $t\geqslant 1$,更有 $f(t)\leqslant f(1)<0$,故 $\lim\limits_{t\to+\infty}f(t)\neq 0$,但由 A 是稳定矩阵又知 $\lim\limits_{t\to+\infty}\mathrm{e}^{At}=0$,$\lim\limits_{t\to+\infty}f(t)=\lim\limits_{t\to+\infty}c^{\mathrm{T}}\cdot\mathrm{e}^{tA^{\mathrm{T}}}\cdot V\cdot\mathrm{e}^{At}\cdot c=0$ 矛盾,故 V 为实正定矩阵.

(2)⇒(1)　令 λ、v 为 A 的特征值及其相应的特征向量,$Av=\lambda v$,V 为矩阵方程 $A^{\mathrm{T}}X+XA=-I$ 的唯一实正定对称解.构造二次型:

$$f(t)=(\mathrm{e}^{\lambda t}v)^*\cdot V\cdot\mathrm{e}^{\lambda t}\cdot v=v^*\cdot\mathrm{e}^{\bar{\lambda}t}\cdot V\cdot\mathrm{e}^{\lambda t}\cdot v=\mathrm{e}^{(\lambda+\bar{\lambda})t}v^*\cdot V\cdot v$$

$$\frac{\mathrm{d}f}{\mathrm{d}t}=(\lambda+\bar{\lambda})\mathrm{e}^{(\lambda+\bar{\lambda})t}v^*\cdot V\cdot v=f(t)\cdot 2\mathrm{Re}(\lambda)$$

但由 $Av=\lambda v$ 知,$(tA)^n v=(t\lambda)^n v$,故又有

$$\begin{aligned}\mathrm{e}^{At}v&=\sum_{n=0}^{\infty}\frac{(At)^n}{n!}v\\&=\sum_{n=0}^{\infty}\frac{(t\lambda)^n}{n!}v\\&=\mathrm{e}^{t\lambda}\cdot v\end{aligned}$$

此时,$f(t)=(\mathrm{e}^{At}v)^*\cdot V\cdot\mathrm{e}^{At}\cdot v=v^*\cdot\mathrm{e}^{tA^{\mathrm{T}}}\cdot V\cdot\mathrm{e}^{At}\cdot v$,再次对 $f(t)$ 求导:

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}t}&=v^*\cdot\mathrm{e}^{tA^{\mathrm{T}}}\cdot A^{\mathrm{T}}\cdot V\cdot\mathrm{e}^{At}v+v^*\cdot\mathrm{e}^{tA^{\mathrm{T}}}V\cdot\mathrm{e}^{At}\cdot Av\\&=v^*\cdot\mathrm{e}^{tA^{\mathrm{T}}}[A^{\mathrm{T}}V+VA]\mathrm{e}^{At}\cdot v\\&=-v^*\cdot\mathrm{e}^{tA^{\mathrm{T}}}\cdot\mathrm{e}^{At}\cdot v\\&=-\mathrm{e}^{\lambda t+\bar{\lambda}t}v^*v\\&=-\mathrm{e}^{(\lambda+\bar{\lambda})t}v^*v\\&=-\mathrm{e}^{2t\mathrm{Re}(\lambda)}v^*v\end{aligned}$$

结果

$$2f(t)\mathrm{Re}(\lambda)=-\mathrm{e}^{2t\mathrm{Re}(\lambda)}v^*v$$

两边让 $t=0$,则得 $2v^*V\cdot v\mathrm{Re}(\lambda)=-v^*v$,由 V 的正定性知:$\mathrm{Re}(\lambda)<0$,这表明 A 为稳定矩阵.

□

习题六

1. 设 $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times n}$是给定矩阵, $X\in\mathbb{R}^{n\times n}$是矩阵变量,标量函数 $\varphi(X)=\mathrm{tr}(AXB)$,证明$\frac{\mathrm{d}\varphi}{\mathrm{d}X}=A^{\mathrm{T}}B^{\mathrm{T}}$.

2. 证明性质 6.1.4 中的等式:$\frac{\mathrm{d}}{\mathrm{d}X}\mathrm{tr}(X^{\mathrm{T}}AX)=(A+A^{\mathrm{T}})X$,特别地,当 A 为对称矩阵时,$\frac{\mathrm{d}}{\mathrm{d}X}\mathrm{tr}(X^{\mathrm{T}}AX)=2AX$.

3. 设 $A_i\in\mathbb{R}^{m\times n}$, $b_i\in\mathbb{R}^{n}$, $i=1,2,\cdots,k$,证明:$x_0\in\mathbb{R}^n$ 是函数$f(x)=\sum_{i=1}^{k}\|A_ix-b_i\|^2$ 的极小值点的充要条件为 $x_0\in\mathbb{R}^n$ 是方程$\left(\sum_{i=1}^{k}A_i^{\mathrm{T}}A_i\right)x=\sum_{i=1}^{k}A_i^{\mathrm{T}}b_i$ 的解.

4. 求 $f=(x-a-Bz)^{\mathrm{T}}(x-a-Bz)$对 a 的导数,其中 x,z 为常向量,B 为常矩阵.

5. 求 $Y^{\mathrm{T}}A^{\mathrm{T}}AY$ 对 A 的导数,其中 $A\in\mathbb{R}^{m\times n}$, $Y\in\mathbb{R}^{m}$ 是常向量.

6. 证明:$\frac{\mathrm{d}(AY-X)^{\mathrm{T}}(AY-X)}{\mathrm{d}A}=2(AY-X)Y^{\mathrm{T}}$,其中 X,Y 为常向量.

7. 若 $A(t)$ 在$[a,b]$上连续可微,证明:$\int_a^t\frac{\mathrm{d}}{\mathrm{d}s}A(s)\mathrm{d}s=A(t)-A(a)$.

8. 设 A 为 2 阶方阵,其两个不同的特征值为 λ_1 和 λ_2,特征向量为 P_1 和 P_2,则方程$\frac{\mathrm{d}x}{\mathrm{d}t}=Ax$ 的解一定能表达为$x(t)=c_1\mathrm{e}^{\lambda_1 t}P_1+c_2\mathrm{e}^{\lambda_2 t}P_2$,其中 c_1 与 c_2 由 $x(0)=c_1P_1+c_2P_2$ 确定.

9. 利用上题结论,求方程$\frac{\mathrm{d}X}{\mathrm{d}t}=\begin{bmatrix}0&1\\-5&-6\end{bmatrix}X$, $X(0)=(1,1)^{\mathrm{T}}$ 的解.

10. 设 $A(t)=\begin{bmatrix}0&\frac{1}{(1+t)^2}\\0&0\end{bmatrix}$,求方程组 $x'(t)=A(t)x(t)$的转移矩阵.

11. 求系统$\begin{pmatrix}x'_1(t)\\x'_2(t)\end{pmatrix}=\begin{bmatrix}0&\frac{1}{t^2+2t+1}\\0&0\end{bmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}1\\1\end{pmatrix}$, $\left.\begin{pmatrix}x_1(t)\\x_2(t)\end{pmatrix}\right|_{t=t_0}=\begin{pmatrix}x_1(t_0)\\x_2(t_0)\end{pmatrix}$的解.

12. 已知 $\Phi(t,t_0)$是方程$\frac{\mathrm{d}x(t)}{\mathrm{d}t}=A(t)X(t)$的转移矩阵,证明:

$$\frac{\mathrm{d}}{\mathrm{d}t_0}\Phi(t,t_0)=-\Phi(t,t_0)A(t_0)$$

13. 求定解问题$\frac{\mathrm{d}X}{\mathrm{d}t}=AX+F(t)$, $X(0)=(1,1,1)^{\mathrm{T}}$ 的解,其中 $A=\begin{bmatrix}3&-1&1\\2&0&-1\\1&-1&2\end{bmatrix}$,

$F(t)=\begin{pmatrix}0\\0\\e^{3t}\end{pmatrix}$.

14. 设某一动态微分方程为：$y'''+7y''+14y'+8y=6u(t)$，其中 y 为系统输出函数，$u(t)$为输入函数，求 $y(t)$.

15. 设 $D=\{X\in\mathbb{R}^{n\times n}:X$ 是正定对称的$\}$，定义 D 上的函数$f(X)=\ln\det(X)$，证明 $\frac{\mathrm{d}f}{\mathrm{d}X}=X^{-1}$.

16. 若矩阵变量 $X\in\mathbb{R}^{n\times n}$ 的行列式 $\det(X)\neq 0$，定义标量函数 $\varphi(X)=\det(X)$，证明 $\frac{\mathrm{d}\varphi}{\mathrm{d}X}=\det(X)(X^{-1})^{\mathrm{T}}$.

第 7 章　特征值估计

科学与工程技术中的许多问题可转化为矩阵的特征值问题．数值矩阵的特征值是一个相似不变量．回顾：任何 n 阶复方阵 A 的谱作为 n 维向量，其欧氏范数不会超过 A 的欧氏范数，它就是 Schur 不等式．那么，对矩阵的特征值，是否还有更精确的估计呢？结论是肯定的．

7.1　特征值界的估计

我们知道，给定 $A=(a_{ij})\in\mathbb{C}^{n\times n}$，总可将 A 进行笛卡儿分解：$A=B+C$，其中 $B=(b_{ij})=\frac{A+A^*}{2}$，$C=(c_{ij})=\frac{A-A^*}{2}$．记 A 的特征值为 $\{\lambda_1,\cdots,\lambda_n\}$，$B$ 的特征值为 $\{\mu_1,\cdots,\mu_n\}$，C 的特征值为 $\{\mathrm{i}v_1,\cdots,\mathrm{i}v_n\}$，并约定

$$|\lambda_1|\geqslant\cdots\geqslant|\lambda_n|,\mu_1\geqslant\cdots\geqslant\mu_n,v_1\geqslant\cdots\geqslant v_n$$

由 Schur 不等式可立即得如下结果：

定理 7.1.1　A,B,C 如上所设，则：

(1) $|\lambda_i|\leqslant n\cdot\max\limits_{i,j}|a_{ij}|=n\|A\|_{v_\infty}$；

(2) $|\mathrm{Re}(\lambda_i)|\leqslant n\cdot\|B\|_{v_\infty}$；

(3) $|\mathrm{Im}(\lambda_i)|\leqslant n\cdot\|C\|_{v_\infty}$．

证明：取 A 的 Schur 分解：$UAU^*=\begin{pmatrix}\lambda_1 & * & * \\ & \ddots & * \\ & & \lambda_n\end{pmatrix}(\triangleq T=(t_{ij}))$，于是

$$UBU^*=\frac{UAU^*}{2}+\frac{UA^*U^*}{2}=\frac{T+T^*}{2}$$

$$UCU^*=\frac{UAU^*}{2}-\frac{UA^*U^*}{2}=\frac{T-T^*}{2}$$

故有

$$\sum_{i=1}^{n}\left|\frac{\lambda_i+\bar{\lambda}_i}{2}\right|+\sum_{j=1}^{n}\sum_{i=1}^{j-1}\frac{|t_{ij}|^2}{2}=\sum_{i,j}|b_{ij}|^2$$

$$\sum_{i=1}^{n}\left|\frac{\lambda_i-\bar{\lambda}_i}{2}\right|+\sum_{j=1}^{n}\sum_{i=1}^{j-1}\frac{|t_{ij}|^2}{2}=\sum_{i,j}|c_{ij}|^2$$

也就是

$$\sum_{i=1}^{n}|\mathrm{Re}(\lambda_i)|^2 \leqslant \sum_{i,j}|b_{ij}|^2 = \|B\|_F^2 \leqslant n^2\max_{i,j}|b_{ij}|^2 = n^2(\|B\|_{v_\infty})^2$$

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2 \leqslant \sum_{i,j}|c_{ij}|^2 = \|C\|_F^2 \leqslant n^2\max_{i,j}|c_{ij}|^2 = n^2(\|C\|_{v_\infty})^2$$

再由 Schur 不等式知

$$\sum_{i=1}^{n}|\lambda_i|^2 \leqslant \sum_{i,j}|a_{ij}|^2 \leqslant n^2\max_{i,j}|a_{ij}|^2 = n^2(\|A\|_{v_\infty})^2$$

结果:

$$|\lambda_i| \leqslant n\cdot\max_{i,j}|a_{ij} = n\|a_{ij}\| = n\|A\|_{v_\infty}$$

$$|\mathrm{Re}(\lambda_i)| \leqslant n\cdot\|B\|_{v_\infty}$$

$$|\mathrm{Im}(\lambda_i)| \leqslant n\cdot\|C\|_{v_\infty}$$

□

由命题 7.2.1 的证明易知如下的 Bendixson 估计式:

推论 7.1.1 若 A 是实 n 阶方阵,则 $|\mathrm{Im}(\lambda_i)| \leqslant \sqrt{\frac{n(n-1)}{2}}\|C\|_{v_\infty}$.

证明: 注意到

$$\sum_{i}|\mathrm{Im}(\lambda_i)|^2 \leqslant \sum_{i,j}|c_{ij}|^2$$

再结合 A 为实方阵,如:$c_{ii}=0,(i=1,2,\cdots,n)$ 且 A 的复特征值共轭成对出现,不妨设为 s 对非实的特征值,于是有

$$2\sum_{i=1}^{s}|\mathrm{Im}(\lambda_i)|^2 \leqslant \sum_{\substack{i,j\\ i\neq j}}|c_{ij}|^2 \leqslant n(n-1)\max_{i,j}|c_{ij}|^2 = n(n-1)(\|C\|_{v_\infty})^2$$

于是得

$$|\mathrm{Im}(\lambda_i)| \leqslant \sqrt{\frac{n(n-1)}{2}}\|C\|_{v_\infty}$$

□

至于 λ_i 与 μ_i, v_i 之间的关系问题,我们有:

定理 7.1.2 $A,B,C,\lambda_i,\mu_i,v_i$ 如前面所设,则:

(1) $\mu_n \leqslant \mathrm{Re}(\lambda_i) \leqslant \mu_1$;

(2) $v_n \leqslant \mathrm{Im}(\lambda_i) \leqslant v_1$.

证明: (1)取相应于 λ_i 的单位特征向量 x,得:

$$\begin{cases} x^*Ax = \lambda_i \\ x^*A^*x = \bar{\lambda}_i \end{cases}$$

于是:

$$\mathrm{Re}(\lambda_i) = \frac{\lambda_i+\bar{\lambda}_i}{2} = x^*\frac{A+A^*}{2}x = x^*Bx$$

$$\mathrm{iIm}(\lambda_i) = \frac{\lambda_i-\bar{\lambda}_i}{2} = x^*\frac{A-A^*}{2}x = x^*Cx$$

注意到 B,C 分别有分解：

$$UBU^* = \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} \triangleq D_1$$

$$VCV^* = \begin{pmatrix} \mathrm{i}v_1 & & \\ & \ddots & \\ & & \mathrm{i}v_n \end{pmatrix} \triangleq D_2$$

并令 $y = Ux, z = Vx$，于是有：

$$\begin{cases} \mathrm{Re}(\lambda_i) = x^* U^* D_1 Ux = y^* D_1 y \\ \mathrm{iIm}(\lambda_i) = x^* V^* D_2 Vx = z^* D_2 z \end{cases}$$

由酉阵不改变向量的 Euclid 范数知，y 和 z 仍是单位向量．从而有

$$\begin{cases} \mathrm{Re}(\lambda_i) = \sum_{i=1}^{n} \mu_i |y_i|^2 \\ \mathrm{iIm}(\lambda_i) = \sum_{j=1}^{n} \mathrm{i}v_j |z_j|^2 \end{cases}$$

亦即

$$\begin{cases} \mathrm{Re}(\lambda_i) = \sum_{i=1}^{n} \mu_i |y_i|^2 \\ \mathrm{Im}(\lambda_i) = \sum_{i=1}^{n} v_j |z_j|^2 \end{cases}$$

根据假定 $\mu_1 \geqslant \cdots \geqslant \mu_n, v_1 \geqslant \cdots \geqslant v_n$，结果：

$$\mu_n = \sum_{i=1}^{n} \mu_n |y_i|^2 \leqslant \sum_{i=1}^{n} \mu_i |y_i|^2 = \mathrm{Re}(\lambda_i) \leqslant \sum_{i=1}^{n} \mu_1 |y_i|^2 = \mu_1$$

同理得 $v_n \leqslant \mathrm{Im}(\lambda_i) \leqslant v_1$.

□

定理 7.1.3 设 $A \in \mathbb{C}^{n\times n}$ 的特征值为 $\lambda_1,\cdots,\lambda_n$，奇异值为 $\sigma_1 \geqslant \cdots \geqslant \sigma_n$，则有

$$\sigma_n \leqslant |\lambda_i| \leqslant \sigma_1, \qquad i = 1,2,\cdots,n$$

证明： 记 $D = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$，则存在酉阵 U，s.t.

$$UA^* AU^* = D$$

取 A 的关于 λ_i 的单位特征向量 x，于是有

$$|\lambda_i|^2 = (Ax)^* Ax = x^* A^* Ax = x^* U^* DUx$$

让 $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = Ux$，于是 y 仍为单位向量，且

$$|\lambda_i|^2 = \sigma_1^2 |y_1|^2 + \cdots + \sigma_n^2 |y_n|^2$$

显然
$$\sigma_n^2 \leqslant |\lambda_i|^2 \leqslant \sigma_1^2$$
亦即
$$\sigma_n \leqslant |\lambda_i| \leqslant \sigma_1$$
□

倘若给定的矩阵 A 是自伴矩阵,则关于其特征值的估计有:

定理 7.1.4(Courant - Fisher 定理)　设 $A \in \mathbb{C}^{n\times n}, A^* = A$. A 的特征值为 $\lambda_1,\cdots,\lambda_n(\lambda_1 \geqslant \cdots \geqslant \lambda_n)$, $R_A(x) = \dfrac{x^* Ax}{x^* x}(0 \neq x \in \mathbb{C}^{n\times 1})$,它称为 A 的 **Rayleigh 商**,则有
$$\lambda_i = \min_{\substack{V \leqslant \mathbb{C}^n \\ \dim V = n-i+1}} \{\max_{0\neq x\in V} R_A(x)\} \quad (i = 1,2,\cdots,n)$$
$$\lambda_i = \max_{\substack{V \leqslant \mathbb{C}^n \\ \dim V = i}} \{\max_{0\neq x\in V} R_A(x)\} \quad (i = 1,2,\cdots,n)$$

证明:记 $\tilde{S} = \{S \mid S为\mathbb{C}^n的 i 维子空间\}$, $\tilde{T} = \{T \mid T为\mathbb{C}^n的 n-i+1 维子空间\}$.将 A 分解为
$$A = U\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}U^*$$
其中 U 为酉矩阵,令 $U = (\alpha_1,\cdots,\alpha_n)$, $U_1 = (\alpha_1,\cdots,\alpha_{i-1})$, $U_2 = (\alpha_i,\cdots,\alpha_n)$,于是 $\mathbb{C}^n = \mathrm{span}\{\alpha_1,\cdots,\alpha_{i-1}\} \oplus \mathrm{span}\{\alpha_i,\cdots,\alpha_n\}$.设 S 是任意 i 维子空间,则 $S \cap \mathrm{span}\{\alpha_i,\cdots,\alpha_n\} \neq 0$,于是可取 $\alpha \in S \cap \mathrm{span}\{\alpha_i,\cdots,\alpha_n\}$ 使得
$$0 \neq \alpha = (\alpha_i,\cdots,\alpha_n)\begin{pmatrix} b_i \\ \vdots \\ b_n \end{pmatrix}$$
此时
$$R_A(\alpha) = \frac{\alpha^* A\alpha}{(\bar{b}_i,\cdots,\bar{b}_n)\begin{pmatrix} b_i \\ \vdots \\ b_n \end{pmatrix}}$$
$$= \frac{(\bar{b}_i,\cdots,\bar{b}_n)\begin{pmatrix} \alpha_i^* \\ \vdots \\ \alpha_n^* \end{pmatrix}(\alpha_1,\cdots,\alpha_n)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}\begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_n^* \end{pmatrix}(\alpha_i,\cdots,\alpha_n)\begin{pmatrix} b_i \\ \vdots \\ b_n \end{pmatrix}}{(\bar{b}_i,\cdots,\bar{b}_n)\begin{pmatrix} b_i \\ \vdots \\ b_n \end{pmatrix}}$$
$$= \frac{\lambda_i |b_i|^2 + \cdots + \lambda_n |b_n|^2}{|b_i|^2 + \cdots + |b_n|^2} \leqslant \lambda_i$$
因此, $\max\limits_{S\in\tilde{S}}\{\min\limits_{0\neq\alpha\in S} R_A(\alpha)\} \leqslant \lambda_i$.

另一方面,取 $S = \mathrm{span}\{\alpha_1,\cdots,\alpha_i\}$,则有 $\dim S = i$,且 $\min\limits_{0\neq\alpha\in S} R_A(\alpha) \geqslant \lambda_i$,因此有

$\max\limits_{S\in\mathcal{S}}\{\min\limits_{0\neq\alpha\in S}R_A(\alpha)\}\geqslant\lambda_i$,结果

$$\lambda_i=\max_{\substack{V\leqslant\mathbb{C}^n\\ \dim V=i}}\{\min_{0\neq x\in V}R_A(x)\}\quad(i=1,2,\cdots,n)$$

类似地可证

$$\lambda_i=\min_{\substack{V\leqslant\mathbb{C}^n\\ \dim V=n-i+1}}\{\max_{0\neq x\in V}R_A(x)\}\quad(i=1,2,\cdots,n)$$

定理获证.

□

上述定理也叫极大极小原理.由该定理立即得下列结果:

命题7.1.1 设 A 为 n 阶 Hermite 阵,其特征值为 $\lambda_1\geqslant\cdots\geqslant\lambda_n$,其相应的正交单位特征向量为 $x_1,\cdots,x_n$. 令 $V_1=\mathbb{C}^n$, $V_i=\{x\in\mathbb{C}^n\mid\langle x,x_k\rangle=0,k=1,2,\cdots,i-1\}$,$(i=2,3,\cdots,n)$.则有

$$\lambda_i=\max_{\substack{x\in V_i\\ x\neq 0}}\frac{\langle Ax,x\rangle}{\langle x,x\rangle},\quad i=1,2,\cdots,n$$

证明: 由于 $\dim V_i=n-i+1$,因此由定理7.1.4知:

$$\max_{\substack{x\in V_i\\ x\neq 0}}\frac{\langle Ax,x\rangle}{\langle x,x\rangle}\geqslant\lambda_i$$

另一方面,$0\neq x\in V_i$,x 可表为:$x=\sum\limits_{t=i}^{n}\langle x,x_t\rangle x_t$,因此有

$$R_A(x)=\frac{\langle Ax,x\rangle}{\langle x,x\rangle}=\frac{\langle\sum\limits_{t=i}^{n}\langle x,x_t\rangle Ax_t,\sum\limits_{t=i}^{n}\langle x,x_t\rangle x_t\rangle}{\sum\limits_{t=i}^{n}|\langle x,x_t\rangle|^2}\leqslant\lambda_i$$

所以 $\lambda_i=\max\limits_{\substack{x\in V_i\\ x\neq 0}}\dfrac{\langle Ax,x\rangle}{\langle x,x\rangle}$, $i=1,2,\cdots,n$.

□

推论7.1.2 若 A,B 均为 n 阶 Hermite 方阵,其特征值分别为:$\lambda_1(A)\geqslant\cdots\geqslant\lambda_n(A)$, $\lambda_1(B)\geqslant\cdots\geqslant\lambda_n(B)$. 则对 $\mathbb{C}^{n\times n}$ 上的任何相容范数 $\|\cdot\|$ 均有 $|\lambda_j(A)-\lambda_j(B)|\leqslant\|A-B\|$, $j=1,2,\cdots,n$.

证明: 由许瓦兹不等式知:

$$|\langle Ax-Bx,x\rangle|\leqslant\|(A-B)x\|_2\|x\|_2\leqslant\|A-B\|_2\|x\|_2^2$$

从而

$$\frac{\langle Ax,x\rangle}{\langle x,x\rangle}\leqslant\frac{\langle Bx,x\rangle}{\langle x,x\rangle}+\frac{\|A-B\|_2\|x\|_2^2}{\langle x,x\rangle}$$

由定理7.1.4知:

$$\lambda_j(A)\leqslant\lambda_j(B)+\|A-B\|_2$$

类似地可得

$$\lambda_j(B)\leqslant\lambda_j(A)+\|A-B\|_2$$

于是

$$|\lambda_j(A)-\lambda_j(B)|\leqslant\|A-B\|_2=\rho(A-B)\leqslant\|A-B\|,j=1,2,\cdots,n$$

这里 $\|A-B\|_2$ 表示算子范数.

□

推论 7.1.3　设 Hermite 阵 $A=(\alpha_{ij})$ 的特征值为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n$,将 $a_{11},\cdots,\alpha_{nn}$ 重排为 $a'_{11},\cdots,a'_{nn}$ 使得 $a'_{11}\geqslant\cdots\geqslant a'_{nn}$.则有

$$|\lambda_i-\alpha'_{ii}|^2\leqslant\sum_{\substack{j,k=1\\j\neq k}}^{n}|a_{jk}|^2,\ i=1,2,\cdots,n$$

证明: 让 $B=\begin{pmatrix}a'_{11}&&\\&\ddots&\\&&a'_{nn}\end{pmatrix}$,将 A 适当交换行和列,使 A 为相似于对角线元依次为 $a'_{11},\cdots,a'_{nn}$ 的矩阵,于是应用推论 7.1.2,并注意到矩阵的下一范数也是相关矩阵范数,得

$$|\lambda_i-a'_{ii}|^2\leqslant\sum_{\substack{j,k=1\\j\neq k}}^{n}|a_{jk}|^2,i=1,2,\cdots,n$$

□

注意到矩阵 A 的奇异值的计算实质上为 A^*A 的特征值计算问题,因此它也有类似于特征值的变分描述.

定理 7.1.5　让 $A\in\mathbb{C}^{m\times n}(m\geqslant n)$ 的奇异值为 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n\geqslant0$. 则有

$$\sigma_i=\min_{\dim S=n-i+1}\max_{\substack{x\in S\\x\neq0}}\frac{\|Ax\|_2}{\|x\|_2}$$

其中变元 S 为$\mathbb{C}^n$ 的子空间 .

证明: 设 σ_i^2 为 A^*A 的第 i 个特征值,因此由极小极大原理知

$$\sigma_i^2=\min_{\dim S=n-i+1}\max_{\substack{x\in S\\x\neq0}}\frac{x^*A^*Ax}{x^*x}$$

$$=\min_{\dim S=n-i+1}\max_{\substack{x\in S\\x\neq0}}\frac{\|Ax\|_2^2}{\|x\|_2^2}$$

结果 $\sigma_i=\min\limits_{\dim S=n-i+1}\max\limits_{\substack{x\in S\\x\neq0}}\dfrac{\|Ax\|_2}{\|x\|_2}$.

□

推论 7.1.4　设 $A,B\in\mathbb{C}^{n\times n}$ 分别具有奇异值 $\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_n$ 及 $\tau_1\geqslant\tau_2\geqslant\cdots\geqslant\tau_n$,则 σ_i 与 τ_i 之间有如下关系:

$$|\sigma_i-\tau_i|\leqslant\|A-B\|_2,\ i=1,2,\cdots,n.$$

证明: 让 $A=B+E$,记 S 为 $n-i+1$ 维子空间,则有

$$\max_{\substack{x\in S\\x\neq0}}\frac{\|Ax\|_2}{\|x\|_2}=\max_{\substack{x\in S\\x\neq0}}\frac{\|Bx+Ex\|_2}{\|x\|_2}$$

$$\leqslant \max_{\substack{x\in S\\x\neq 0}} \frac{\| Bx \|_2}{\| x \|_2} + \max_{\substack{x\in S\\x\neq 0}} \frac{\| Ex \|_2}{\| x \|_2}$$

$$\leqslant \max_{\substack{x\in S\\x\neq 0}} \frac{\| Bx \|_2}{\| x \|_2} + \| A - B \|_2$$

从而有 $\sigma_i - \tau_i \leqslant \| A - B \|_2$.类似地,有 $\tau_i - \sigma_i \leqslant \| A - B \|_2$. 于是有 $|\sigma_i - \tau_i| \leqslant \| A - B \|_2, i = 1,2,\cdots,n$.证毕.

□

7.2 戈氏圆盘定理

现在我们将推论 7.1.3 推广到一般矩阵中,从而给出特征值在 $\mathbb{C}$ 上的位置的初步估计.

定理 7.2.1 设 $A = (a_{ij})$ 为复 n 阶方阵,定义圆盘

$$G_j(A) = \{\lambda \in C : |\lambda - a_{jj}| \leqslant \sum_{\substack{k=1\\k\neq j}}^{n} |a_{jk}|\},\ j = 1,2,\cdots,n$$

而
$$G_j^*(A) = \{\lambda \in C : |\lambda - a_{jj}| \leqslant \sum_{\substack{k=1\\k\neq j}}^{n} |a_{kj}|\},\ j = 1,2,\cdots,n$$

则 A 的特征值 λ 必满足:

$$\lambda \in (\bigcup_{j=1}^{n} G_j) \cap (\bigcup_{j=1}^{n} G_j^*)$$

这里 G_j 是 $G_j(A)$ 的简记,G_j^* 是 $G_j^*(A)$ 的简记.以下约定用 δ_j 表圆盘 $G_j(A)$ 的半径.

证明: 假定 $Ax = \lambda x$, $\| x \|_\infty = 1$,记 $x = (x_1,\cdots,x_n)^{\mathrm{T}}$,假定 $|x_j| = 1$,则有

$$|\lambda - a_{jj}| = |(\lambda - a_{jj})x_j| = \left|\sum_{\substack{k=1\\k\neq j}}^{n} a_{jk}x_k\right| \leqslant \sum_{\substack{k=1\\k\neq j}}^{n} |a_{jk}|$$

即有 $\lambda \in \bigcup_{j=1}^{n} G_j$. 由于 A^{T} 与 A 有相同的特征值,因此本定理得证.

□

本定理连同下列定理7.2.2统称为 **Gersgorin 圆盘定理**,它是由 Gersgorin 于 1931 年获得的. $G_j(A)$ 叫做 A 的一个 Gersgorin 圆盘(简称**戈氏圆盘**).

定理 7.2.2 设 $\{i_1,\cdots,i_n\}$ 为 $\{1,2,\cdots,n\}$ 的一个排列.

$$S = \bigcup_{j=1}^{s} G_{i_j}(A),\ T = \bigcup_{j=s+1}^{n} G_{i_j}(A)$$

若 $S \cap T = \varnothing$,则 S 中必含有 A 的特征值.

为证明此定理,我们先给出以下几个引理.

引理 7.2.1 若 $\lim\limits_{m\to\infty} A_m = A$,则 λ 为 A 的特征值当且仅当存在由 A_m 的特征值 λ_m 构成的点列 $\{\lambda_m\}$ s.t. $\lim\limits_{m\to\infty} \lambda_m = \lambda$.

证明： 让 n 阶方阵A_m 的特征多项式为$\chi_m(x)$，n 阶方阵A 的特征多项式为$\chi(x)$. 因此有

$$\lim_m \det(xI - A_m) = \det(xI - A)$$

即有

(1)$\lim_m \chi_m(x) = \chi(x)$，对所有 x；

(2)$\chi_m(x) = \prod_{j=1}^{n}(x - \lambda_j^{(m)})$，其中$\{\lambda_1^{(m)}, \cdots, \lambda_n^{(m)}\}$ 为 A_m 的谱；$\chi(x) = \prod_{j=1}^{n}(x - \lambda_j)$，其中$\{\lambda_1, \cdots, \lambda_n\}$ 为 A 的谱 .

设 λ 为A 的特征值，则由(1)知$\lim_m \chi_m(\lambda) = 0$，从而任给定正整数 k，$\exists$ 正整数 N_k，使得当 $m \geqslant N_k$ 时有$|\chi_m(\lambda)| < \left(\frac{1}{k}\right)^n$，从而应用(2)我们知：若 $m \geqslant N_k$，则对每个 m，A_m 中必有一个特征值 $\lambda_{j_k}^{(m)}$，使得$\left|\lambda - \lambda_{j_k}^{(m)}\right| < \frac{1}{k}$. 不失一般性，假定 $N_1 < N_2 < \cdots$. 若 $m \geqslant N_1$，则必有对某个 k 使得$N_k \leqslant m < N_{k+1}$，于是取$\lambda_m = \lambda_{j_k}^{(m)}$；若 $m < N_1$，我们取$\lambda_m = \lambda_1^{(m)}$，这样得到一个数列$\{\lambda_m\}_{m=1}^{\infty}$. 显然 $\forall \varepsilon > 0$，可取正整数 p，s.t. $0 < \frac{1}{p} \leqslant \varepsilon$. 于是若 $m \geqslant N_p$，则有 $N_p \leqslant N_k \leqslant m < N_{k+1}$，从而

$$|\lambda - \lambda_m| = \left|\lambda - \lambda_{j_k}^{(m)}\right| < \frac{1}{k} \leqslant \frac{1}{p} \leqslant \varepsilon$$

结果$\lim_m \lambda_m = \lambda$.

反过来，若有 A_m 的特征值λ_m 构成的收敛于λ 的数列，将 $\chi_m(x)$、$\chi(x)$ 展开为多项式表示：

$$\chi_m(x) = \sum_{j=0}^{n} a_j^{(m)} x^j, \chi(x) = \sum_{j=0}^{n} a_j x^j$$

于是

$$|\chi_m(x) - \chi(x)| \leqslant \sum_{j=1}^{n} |a_j^{(m)} - a_j| |x^j|，对所有 x 成立$$

由于$\lim_m \lambda_m = \lambda$，因此 $\exists \rho \geqslant 1$ 使得$|\lambda| \leqslant \rho$，$|\lambda_m| \leqslant \rho$ 对所有 m 等成立，故有：

$$|\chi_m(\lambda_m) - \chi(\lambda_m)| \leqslant \rho^n\left(\sum_{j=1}^{n} |a_j^{(m)} - a_j|\right)对所有 m 等成立$$

由$\lim_m A_m = A$ 知：$\lim_m a_j^{(m)} = a_j (1 \leqslant j \leqslant n)$，进一步有 $\chi_m(\lambda_m) = 0$，结果

$$\lim_m \chi(\lambda_m) = 0$$

即有 $\chi(\lambda) = 0$，λ 为A 的特征值，引理证毕 .

□

本引理的一个有用的推论是：

推论 7.2.1 设$\lim_m A_m = A$，则有$\lim_m \rho(A_m) = \rho(A)$.

证明：设 λ 为 A 的特征值满足 $|\lambda| = \rho(A)$. 于是 $\exists\{\lambda_m\}$，λ_m 为 A_m 的特征值，满足 $\lim\limits_m \lambda_m = \lambda$. 从而

$$\underline{\lim_m}\rho(A_m) \geqslant \lim_m |\lambda_m| = \rho(A)$$

另一方面，由 $\lim\limits_m A_m = A$ 知 $\lim\limits_m \|A_m\| = \|A\|$，其中 $\|\ \|$ 为任意相容的矩阵范数. 结果

$$\rho(A_m) \leqslant \|A_m\|, \overline{\lim_m}\rho(A_m) \leqslant \lim \|A_m\| = \|A\|$$

由 $\|\ \|$ 的任意性得 $\overline{\lim}\rho(A_m) \leqslant \rho(A)$.

□

定理 7.2.2 的证明　不妨设 $G_{i_j}(A)$ 的半径为 $\delta_{i_j} > 0$，其中 $1 \leqslant j \leqslant s$，定义 $A(\tau) = \begin{pmatrix} a_{11} & \tau a_{12} & \cdots & \tau a_{1,n-1} & \tau a_{1n} \\ \vdots & \vdots & & \vdots & \vdots \\ \tau a_{n1} & \tau a_{n2} & \cdots & \tau a_{n,n-1} & a_{nn} \end{pmatrix}$，$\tau \in [0,1]$. $A(\tau)$ 的特征值集合为 $\Lambda(\tau)$，则有 $\Lambda(0) = \{a_{i_j i_j} : 1 \leqslant j \leqslant n\}$.

(1) $[0,1]$ 中必存在某 σ，使得对所有的 $0 < \tau \leqslant \sigma$ 均有 $\Lambda(\tau) \cap S \neq \varnothing$. 否则，对每个 $m \in \mathbb{N}$ 必有一个 τ_m s.t.

① $\Lambda(\tau_m) \cap S = \varnothing$;

② $0 < \tau_m \leqslant \dfrac{1}{m}$.

此时，$\lim\limits_m A(\tau_m) = A(0)$，从而由引理 7.2.1 知，

$$\exists \lambda^{(m)} \in \Lambda(\tau_m) \text{ 使得} \lim_m \lambda^{(m)} = a_{i_1 i_1}$$

继而 m 充分大后有 $|a_{i_1 i_1} - \lambda^{(m)}| < \delta_{i_1}$. 换句话说，$m$ 充分大后 $\lambda^{(m)} \in G_{i_1}(A)$，矛盾于 $\Lambda(\tau_m) \cap S = \varnothing$.

现令 $G = \{\alpha : 0 < \alpha \leqslant 1$，使得 $\Lambda(\tau) \cap S \neq \varnothing$ 对所有满足 $0 < \tau \leqslant \alpha$ 的 τ 皆成立$\}$，再取 G 的上确界为 μ，则显见有下列结论成立：

(2) $\Lambda(\tau) \cap S \neq \varnothing$ 对一切满足 $0 < \tau < \mu$ 的 τ 皆成立.

所以当 m 充分大后，$\Lambda(\mu - \dfrac{1}{m}) \cap S \neq \varnothing$. 注意到 S 为紧集，因此有一个特征值系列 λ_{m_k}（其中 λ_{m_k} 为 $A(\mu - \dfrac{1}{m_k})$ 的特征值）收敛于 S 中的某点 λ. 再次由引理 7.2.1 知，该 λ 为 $A(\mu)$ 的特征值. 进而知(3).

(3) $\Lambda(\mu) \cap S \neq \varnothing$.

现若 $\mu < 1$，则取 m 充分大后有 $\mu + \dfrac{1}{m} < 1$. 于是 $\exists \sigma_m \in \left(\mu, \mu + \dfrac{1}{m}\right]$ 使得 $\Lambda(\sigma_m) \cap S = \varnothing$. 这样，我们在 $(\mu, 1)$ 中获得收敛于 μ 的一个无限点列 $\{\sigma_m\}$，它满足 $\Lambda(\sigma_m) \cap S = \varnothing$，从而 $\Lambda(\sigma_m) \subseteq T$. 注意 T 也为有界闭集，结果由 $\lim\limits_m A(\sigma_m) = A(\mu)$ 结合引理 7.2.1 知：

$\Lambda(\mu) \subseteq T$,矛盾于 $S \cap T = \emptyset$.因此 μ 只能为1,从而 S 包含 A 的一个特征值.定理7.2.2获证.

□

推论 7.2.2　对戈氏圆盘两两互不相交的实矩阵,其特征值均是实的,且可对角化.

证明: 由于 n 个戈氏圆盘两两不相交,因此有 n 个不同特征值,故可对角化.进一步,实矩阵的戈氏圆盘皆是以实轴为对称中心,而实矩阵的特征根要么为实的,要么共轭成对出现,因此,这 n 个不同特征值只能均是实的.

□

作为戈氏圆盘定理的应用,我们还可直接得到如下 Ky Fan 定理.

定理 7.2.3　若 A 为 n 阶复方阵,$B = (b_{ij}), b_{ij} \geqslant |a_{ij}|, i,j = 1,2,\cdots,n, \lambda$ 为 A 的一个特征值.则 $\exists\, i \in \{1,2,\cdots,n\}$ 使得 $|\lambda - a_{ii}| \leqslant \rho(B) - b_{ii}$.

证明: B 为非负矩阵,不妨设 $B > 0$,于是由 Perron 定理可知:$\exists$ 正向量 v, s.t. $Bv = \rho(B)v$. 记 $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$,构作矩阵 $V = \begin{pmatrix} v_1 & & \\ & \ddots & \\ & & v_n \end{pmatrix}$,令 $C = V^{-1}AV$,则有 $c_{ii} = a_{ii}$, $(i = 1, 2,\cdots,n)$,且对某个 i,有

$$|\lambda - c_{ii}| \leqslant \sum_{j \neq i} |c_{ij}|$$

进而

$$\begin{aligned} |\lambda - a_{ii}| &\leqslant \sum_{j\neq i} |v_i^{-1} a_{ij} \cdot v_j| = \sum_{j \neq i} |v_j \cdot v_i^{-1} \cdot a_{ij}| \\ &\leqslant \sum_{j\neq i} |v_j \cdot v_i^{-1} \cdot b_{ij}| = v_i^{-1}\Big(\sum_{j\neq i} |v_j \cdot b_{ij}|\Big) \\ &= v_i^{-1}\Big(\sum_{j=1}^{n} b_{ij}v_j - b_{ii}v_i\Big) = v_i^{-1}(\rho(B)v_i - b_{ii}v_i) \\ &= \rho(B) - b_{ii} \end{aligned}$$

当 B 不是正矩阵时,让 $b_{ij}^{(m)} = b_{ij} + \dfrac{1}{m}$ 对所有 (i,j),则有 $B^{(m)} \triangleq (b_{ij}^{(m)})$ 为正矩阵.对 $B^{(m)}$,有某 $i(m) \in \{1,2,\cdots,n\}$ 使得

$$|\lambda - a_{i(m)}i(m)| \leqslant \rho(B^{(m)}) - \Big(b_{i(m)}i(m) + \frac{1}{m}\Big)$$

再由两边取极限,则由 $\lim\rho(B^{(m)}) = \rho(B)$ 立即知本定理获证.

□

定理 7.2.4(Ostrowski 定理)　设 A 为 n 阶复方阵,$0 \leqslant \alpha \leqslant 1$. λ 为 A 的特征值,则 $\exists\, i \in (1 \leqslant i \leqslant n)$ 使得

$$|\lambda - a_{ii}| \leqslant \delta_i^{\alpha}(A) \cdot \delta_i^{1-\alpha}(A^{\mathrm{T}})$$

其中 $\delta_i(A)$ 表 A 的第 i 个戈氏圆盘的半径,A^{T} 表示 A 的转置.

证明: 当 $\alpha = 0$ 或 $\alpha = 1$ 时,欲证的定理为戈氏圆盘定理.

假定 $\alpha \in (0,1)$, $(\lambda I - A)x = 0$ 对某个 $x \neq 0$,则对所有的 $i \leqslant n$ 均有:

$$|\lambda - a_{ii}||x_i| = \left|\sum_{j\neq i} a_{ij}x_j\right| \leqslant \sum_{j\neq i}|a_{ikj}||x_j|$$

反设定理 7.2.4 不成立，让 $\Omega_i = \delta_i^{\alpha}(A)\cdot\delta_i^{1-\alpha}(A^{\mathrm{T}})$，则对 $i = 1,2,\cdots,n$ 均有 $\Omega_i < |\lambda - a_{ii}|$. 由 $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \neq 0$，取集合 $J = \{i \leqslant n : x_i \neq 0\}$，则 $J \neq \varnothing$，且 $\forall i \in J$ 有

$$\Omega_i|x_i| < |\lambda - a_{ii}||x_i| \leqslant \sum_{j\neq i}|a_{ij}|^{\alpha}|a_{ij}|^{1-\alpha}|x_j|$$

利用 Hölder 不等式知

$$\begin{aligned}\forall i \in J, \Omega_i|x_i| &< \sum_{j\neq i}|a_{ij}|^{\alpha}(|a_{ij}|^{1-\alpha}|x_j|)\\ &\leqslant \Big(\sum_{j\neq i}|a_{ij}|^{\alpha/\alpha}\Big)^{\alpha}\cdot\Big(\sum_{j\neq i}|a_{ij}|^{(1-\alpha)/(1-\alpha)}|x_j|^{1/(1-\alpha)}\Big)^{1-\alpha}\\ &= \Big(\sum_{j\neq i}|a_{ij}|\Big)^{\alpha}\cdot\Big(\sum_{j\neq i}|a_{ij}||x_j|^{1/(1-\alpha)}\Big)^{1-\alpha}\end{aligned}$$

倘若 $\delta_i^{\alpha}(A) = 0$，则上述表明 $0 < 0$，矛盾.

倘若 $\delta_i^{\alpha}(A) \neq 0$，则 $\delta_i^{1-\alpha}(A^{\mathrm{T}})|x_i| < \Big(\sum_{j\neq i}|a_{ij}||x_j|^{1/(1-\alpha)}\Big)^{1-\alpha}$.

两边开 $1-\alpha$ 次方得，$\delta_i(A^{\mathrm{T}})|x_i|^{1/(1-\alpha)} < \sum_{j\neq i}|a_{ij}||x_j|^{1/(1-\alpha)}$，进而得

$$\begin{aligned}\sum_{i=1}^{n}\delta_i(A^{\mathrm{T}})|x_i|^{1/(1-\alpha)} &< \sum_{i=1}^{n}\sum_{j\neq i}|a_{ij}||x_j|^{1/(1-\alpha)}\\ &= \sum_{i=j}^{n}\sum_{i\neq j}|a_{ij}||x_j|^{1/(1-\alpha)}\\ &= \sum_{j=1}^{n}\Big(|x_j|^{1/(1-\alpha)}\sum_{i\neq j}|a_{ij}|\Big)\\ &= \sum_{j=1}^{n}\big(|x_j|^{1/(1-\alpha)}\delta_j(A^{\mathrm{T}})\big)\\ &= \sum_{i=1}^{n}\delta_i(A^{\mathrm{T}})|x_i|^{1/(1-\alpha)}\end{aligned}$$

矛盾.

□

利用 Ostrowski 定理，立即得：

推论 7.2.3 设 A 为 n 阶复方阵，$\alpha \in [0,1]$，则

(1) 当 A 为奇异阵时，$\exists\, i \in (1 \leqslant i \leqslant n)$ 使 $|a_{ii}| \leqslant \delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^{\mathrm{T}})$；

(2) 当 A 为奇异阵时，$\exists\, i \in (1 \leqslant i \leqslant n)$ 使 $|a_{ii}| \leqslant \alpha\delta_i(A) + (1-\alpha)\delta_i(A^{\mathrm{T}})$；

(3) $\rho(A) \leqslant \max_i\{|a_{ii}| + \delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^{\mathrm{T}})\}$；

(4) $\rho(A) \leqslant \max_i\{|a_{ii}| + \alpha\delta_i(A) + (1-\alpha)\delta_i(A^{\mathrm{T}})\}$；

(5) 记 $R_i(A) = \sum_{j=1}^{n}|a_{ij}|$，则有 $\rho(A) \leqslant \max_i\{R_i^{\alpha}(A)R_i^{1-\alpha}(A^{\mathrm{T}})\}$.

证明：这里只证(5)，其余均是显然的．由本推论中的(3)有

$$\rho(A) \leqslant \max_i \{ |a_{ii}| + \delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^T) \}$$

注意到$|a_{ii}| + \delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^T) = |a_{ii}|^{\alpha} \cdot |a_{ii}|^{1-\alpha} + \delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^T)$,同时由 Hölder 不等式知:

$$\begin{aligned}|a_{ii}| + \delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^T) &\leqslant [(|a_{ii}|^{\alpha})^{1/\alpha} + (\delta_i^{\alpha}(A))^{1/\alpha}]^{\alpha} \\ &\quad \cdot [(|a_{ii}|^{1-\alpha})^{1/(1-\alpha)} + (\delta_i^{1-\alpha}(A^T))^{1/(1-\alpha)}]^{1-\alpha} \\ &= [|a_{ii}| + \delta_i(A)]^{\alpha} \cdot [|a_{ii}| + \delta_i(A^T)]^{1-\alpha} \\ &= R_i^{\alpha}(A) \cdot R_i^{1-\alpha}(A^T)\end{aligned}$$

□

Ostrowski 还以另外一种方式推广了 Gersgorin 圆盘定理．对 n 阶复方阵A,记

$$O_{ij}(A) = \{z \in \mathbb{C} : |z - a_{ii}| \cdot |z - a_{jj}| \leqslant \delta_i(A)\delta_j(A)\}$$

对任何 $i \neq j$.

定理 7.2.5　设 $n \geqslant 2$,则 n 阶复方阵A 的特征值λ 必属于某个O_{ij}.

证明: $(\lambda I - A)x = 0, x \neq 0$. 记 $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$,选取 r 和t 使得$r \neq t$,且$|x_r| \geqslant |x_t| \geqslant |x_j|$,其中 $j \in \{1,2,\cdots,n\} \setminus \{r,t\}$.

若 $x_t = 0$,于是有 $x_1,\cdots,x_n$ 中除x_r 外均为 0,结果

$$\lambda x_r = \sum_{j=1}^{n} a_{rj}x_j = a_{rr}x_r, \lambda = a_{rr}$$

结论真.若 $x_t \neq 0$,则对所有 $i \leqslant n$ 有

$$(\lambda - a_{ii})x_i = \sum_{j \neq i} a_{ij}x_j$$

特别地,有

$$\begin{aligned}|(\lambda - a_{rr})||x_r| &= \left|\sum_{j \neq r} a_{rj}x_j\right| \leqslant \sum_{j \neq r} |a_{rj}| \cdot |x_j| \leqslant |x_t| \cdot \delta_r(A) \\ |(\lambda - a_{tt})||x_t| &= \left|\sum_{j \neq t} a_{tj}x_j\right| \leqslant \sum_{j \neq t} |a_{tj}| \cdot |x_j| \leqslant |x_r| \cdot \delta_t(A)\end{aligned}$$

结果$|(\lambda - a_{rr})||(\lambda - a_{tt})||x_r||x_t| \leqslant |x_t \cdot x_r| \cdot \delta_r(A) \cdot \delta_t(A)$、获证.

□

推论 7.2.4　$n \geqslant 2$,对任何 n 阶复方阵$A = (a_{ij})$,若 $\forall i \neq j$ 皆有$|a_{ii}| \cdot |a_{jj}| > \delta_i(A) \cdot \delta_j(A)$,则 A 必可逆.

Ostrowski 定理还可扩展为如下定理.

定理 7.2.6　设 A 为n 阶复方阵($n > 1$),λ 为A 的一个特征值,$0 \leqslant \alpha \leqslant 1$,则 $\exists i \neq j$,使得

$$|\lambda - a_{ii}||\lambda - a_{jj}| \leqslant [\delta_i(A) \cdot \delta_j(A)]^{\alpha}[\delta_i(A^T) \cdot \delta_j(A^T)]^{1-\alpha}$$

证明: 先证如下断言.

断言: 设 $A = (a_{ij})_{n \times n} \in \mathbb{C}^{n \times n}, \alpha \in [0,1]$,且对所有的 $i \neq j$,满足

$$|a_{ii}||a_{jj}| > [\delta_i(A)\cdot\delta_j(A)]^{\alpha}[\delta_i(A)^{\mathrm{T}}\cdot\delta_j(A)^{\mathrm{T}}]^{1-\alpha}$$

则必有 A 可逆.

事实上,在断言假设条件下,令 $s_i(A)=\delta_i^{\alpha}(A)\delta_i^{1-\alpha}(A^{\mathrm{T}})/|a_{ii}|$ 得 $\{s_1(A),\cdots,s_n(A)\}$. 重新排序使 $s_{m_1}(A)\geqslant\cdots\geqslant s_{m_n}(A)$,于是由假设条件有:对一切 $i\neq j$, $s_i(A)s_j(A)<1$. 特别地, $s_{m_1}(A)s_{m_2}(A)<1$,从而 $s_{m_2}(A)<1$.

Case 1 $s_{m_2}(A)=0$,此时 $s_{m_3}(A)=\cdots=s_{m_n}(A)=0$,结果容易计算得 $\det A=a_{11}\cdots a_{nn}\neq 0$;

Case 2 $s_{m_2}(A)\neq 0$,此时 $s_{m_2}(A)\in(0,1)$,令 $q=\sqrt{\dfrac{s_{m_1}(A)}{s_{m_2}(A)}}$ 并让 B 为 A 的第 m_1 行和第 m_1 列均乘以 q 后得到的矩阵,比较 A 与 B 有:

$$\delta_{m_1}(B)=q\delta_{m_1}(A),\ \delta_{m_1}(B^{\mathrm{T}})=q\delta_{m_1}(A^{\mathrm{T}}),\ |b_{m_1m_1}|=q^2|a_{m_1m_1}|$$

因而有 $s_{m_1}(B)=\dfrac{s_{m_1}(A)}{q}=\sqrt{s_{m_1}(A)s_{m_2}(A)}<1$. 而且对 $i=2,3,\cdots,n$ 时有

$$s_{m_i}(B)\leqslant qs_{m_i}(A)\leqslant\sqrt{s_{m_1}(A)s_{m_2}(A)}<1$$

再次由 $s_{m_i}(B)$ 的定义得

$$|b_{ii}|>\delta_i^{\alpha}(B)\delta_i^{1-\alpha}(B^{\mathrm{T}}),\ i=1,2,\cdots,n$$

结果由定理 7.2.4 知 B 可逆,注意到 $\det B=q^2\det A$,故 A 也可逆,断言真.

现在回到本定理的证明.反证,若对所有的 $i\neq j$ 均有

$$|(\lambda-a_{ii})||(\lambda-a_{jj})|>[\delta_i(A)\cdot\delta_j(A)]^{\alpha}[\delta_i(A^{\mathrm{T}})\cdot\delta_j(A^{\mathrm{T}})]^{1-\alpha}$$

则由断言知: $\det(\lambda I-A)\neq 0$. 这与 λ 为 A 的特征值相矛盾.证毕.

□

戈氏圆盘定理在确定特征值的位置,甚至分离特征值等方面有着重要的意义.

例 7.2.1 让 $A=\begin{pmatrix}2 & 1 & 0\\ \frac{1}{5} & 3 & \frac{1}{4}\\ \frac{1}{6} & \frac{1}{4} & 6\end{pmatrix}$,判定 A 是否可对角化?

解: 让 $P=\begin{pmatrix}\frac{1}{2} & & \\ & 1 & \\ & & \frac{10}{3}\end{pmatrix}$,于是 $PAP^{-1}=\begin{pmatrix}2 & \frac{1}{2} & 0\\ \frac{2}{5} & 3 & \frac{0.3}{4}\\ \frac{10}{9} & \frac{5}{6} & 6\end{pmatrix}$. PAP^{-1} 的戈式圆盘为:

$$G_1=\{z:|z-2|\leqslant\frac{1}{2}\}$$

$$G_2=\{z:|z-3|\leqslant 0.475\}$$

$$G_3 = \{z : |z-6| \leqslant \frac{35}{18}\}$$

由于 G_1, G_2, G_3 互不相交，所以 A 有三个不同的特征值，可实对角化.

□

习题七

1. 设 $A \in \mathbb{R}^{n\times n}$，且 $|a_{ii}| > \sum_{j\neq i} |a_{ij}| (i = 1,2,\cdots,n)$. 则 $\det A$ 与 $\prod_{i=1}^{n} a_{ii}$ 同号.

2. 设 $A \in \mathbb{R}^{n\times n}$ 满足 $a_{ii} > \sum_{j\neq 1} |a_{ij}| (i = 1,2,\cdots,n)$. 则 A 的所有特征值均位于复平面的右半平面.

3. 设 $A \in \mathbb{C}^{n\times n}$，记 $R_i(A) = \sum_{j=1}^{n} |a_{ij}|$. 证明：

(1) $\rho(A) \leqslant \max_i \sqrt{R_i(A) R_i(A^{\mathrm{T}})}$；　　(2) $\rho(A) \leqslant \min_i \{\|A\|_\infty, \|A\|_1\}$；

(3) $\rho(A) \leqslant \max_i \frac{R_i(A) + R_i(A^{\mathrm{T}})}{2}$；　　(4) $\rho(A) \leqslant \frac{1}{2}(\|A\|_\infty + \|A\|_1)$.

4. 设 $A \in \mathbb{C}^{n\times n}$，$c_1, c_2, \cdots, c_n$ 为任意一组正数，令 $K_r = \frac{\sum_{i=1}^{n} c_i |a_{ri}|}{c_r}$，$r = 1,2,\cdots,n$，$K = \max_r K_r$. 证明：$\rho(A) \leqslant K$.

5. 设 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$，$\alpha \in [0,1]$ 且 $|a_{ii}| > \delta_1^{\alpha}(A)\delta_i^{1-\alpha}(A^{\mathrm{T}}) (i = 1,2,\cdots,n)$. 证明：$A$ 可逆.

6. 设 A 是非负矩阵，证明：$r \leqslant \rho(A) \leqslant R$，其中 $R = \|A\|_\infty$，$r = \min_i \{\sum_{j=1}^{n} a_{ij}\}$.

7. 设 $A = (a_{ij})$ 为正矩阵，且 $R_i = \sum_{j=1}^{n} a_{ij}$，$r = \min_i R_i$，$R = \max_i R_i = \|A\|_\infty$，$m = \min_{i,j} a_{ij}$，$\delta = \max_{R_i < R_j} \frac{R_i}{R_j}$. 证明：$r < R$ 时有 $r + m(h-1) \leqslant \rho(A) \leqslant R - m(1 - g^{-1})$，其中：

$$g = \frac{[R - 2m + \sqrt{R^2 - 4m(R-r)}]}{2(r-m)}$$

$$h = \frac{[-r + 2m + \sqrt{r^2 + 4m(R-r)}]}{2m}$$

8. 设 $A \in \mathbb{R}^{n\times n}$ 的所有非主对角元 $a_{ij} \geqslant 0$，且存在一组正数 $c_1, \cdots, c_n$ 使得

$$\sum_{j=1}^{n} a_{ij} c_j \leqslant 0, i = 1,2,\cdots,n$$

证明：A 是半稳定矩阵.

9. 设 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$. 若存在一组正数 $k_1, k_2, \cdots, k_n$，使得

$$\sum_{\substack{j=1\\j\neq i}}^{n} k_j |a_{ij}| < -k_i \mathrm{Re}(a_{ii}), i = 1,2,\cdots,n$$

则 A 是稳定矩阵.

10. 设 $A \in \mathbb{R}^{n\times n}$ 满足 $a_{ii} \leqslant -\sum_{\substack{j=1\\j\neq 1}}^{n} |a_{ij}|, i = 1,2,\cdots,n$.证明:$A$ 是稳定的当且仅当 A 为可逆阵.

11. 设 $B \in \mathbb{C}^{n\times n}$ 为可对角化矩阵,$\rho(B) < 1$.证明:∃ 正定阵 A 使得 $A - B^* AB$ 仍为正定.

12. 设 A 与 $A - 4B^* AB$ 均为 Hermite 正定阵.证明:$\rho(B) < \frac{1}{2}$.

13. 设 $A \in \mathbb{C}^{n\times n}$,证明:$\lambda_i\left(\frac{A^* + A}{2}\right) \leqslant \sigma_i(A) = \sqrt{\lambda_i(A^* A)}$,其中 $\lambda_1\left(\frac{A^* + A}{2}\right) \leqslant \cdots \leqslant \lambda_n\left(\frac{A^* + A}{2}\right)$ 为 $\left(\frac{A^* + A}{2}\right)$ 的 n 个特征值,$\lambda_1(A^* A) \leqslant \cdots \leqslant \lambda_n(A^* A)$ 为 $A^* A$ 的 n 个特征值.

14. 证明矩阵 $A = \begin{pmatrix} 1 & n^{-1} & n^{-1} & \cdots & n^{-1} \\ n^{-1} & 3 & n^{-1} & \cdots & n^{-1} \\ \vdots & \vdots & \vdots & & \vdots \\ n^{-1} & n^{-1} & n^{-1} & \cdots & 2n-1 \end{pmatrix}_{n\times n}$ 是可实对角化的,并且 A 的特征值均是实数.求出秩 A.

参考文献

[1] 陈公宁.矩阵理论与应用.北京:科学出版社,2007

[2] 吴昌悫.魏洪增.矩阵理论与方法.北京:电子工业出版社,2006

[3] 苏育才,姜翠波,张跃辉.矩阵理论.北京:科学出版社,2006

[4] 戴华.矩阵论.北京:科学出版社,2005

[5] (美)合恩(Horn R A),等.矩阵分析.杨奇,译.北京:机械工业出版社,2005

[6] 黄有度,狄成恩,朱士信.矩阵论及其应用.合肥:中国科学技术大学出版社,2005

[7] 陈祖明,等.矩阵论引论.北京:北京航空航天大学出版社,1998

[8] 史荣昌.矩阵分析.北京:北京理工大学出版社,2004

[9] 罗家洪.矩阵分析引论.北京:华南理工大学出版社,1996

[10] 王朝瑞,等.矩阵分析.北京:北京理工大学出版社,1987

[11] 程云鹏,等.矩阵论.西安:西北工业大学出版社,2006

[12] 北京大学几何与代数教研室编写组.高等代数.北京:人民教育出版社,1978

[13] 南京大学计算数学专业编写组.线性代数.北京:科学出版社,1978

[14] 陈景良,陈向辉.特殊矩阵.北京:清华大学出版社,2001

[15] 张贤达.矩阵分析与应用.北京:清华大学出版社,2004

[16] Lancaster P, Tismenetsky M. The Theory of Matrices with Applications, 2nd Edition. Academic Press, 1985

[17] Barth T, Manteuffel T. Multiple Recursion Conjugate Gradient Algorithms Part I: Sufficient Condition. SIAM J. Matrix Anal. Appl., 2000

[18] Beattie C, Fox D. Schur Complements and the Weinstein-Aronszajn Theory for Modified Matrix Eigenvalue Problems. LAA, 1988

[19] Bellman R. Introduction to Matrix Analysis, 2nd edition. McGraw-Hill, 1970

[20] Champagne B. Adaptive Eigendecomposition of Data Covariance Matrices Based on First-order Perturbations. IEEE Trans Signal Processing, 1994

[21] Chatelin F. Eigenvalues of Matrices. Wiley, 1993

[22] De Moor B. Total Least Squares for Affinety Structured Matrices and the Noisy Realization Problem. IEEE Trans Signal Processing, 1994

[23] Gilllies A W. On the Classification of Matrix Generalized Inverse. SIAM Review, 1970

[24] Gantmacher F R. The theory of Matrices. Chelsea Publishing, 1977

[25] Cantmacher F R. Applications of the Theory of Matrices. Interscience, 1959

[26] Holmgren S, Otto K. Iterative Solution Methods and Preconditioners for Block-tridiagonal Systems of Equations. SIAM J. Matrix Anal. Appl., 1992

[27] Henderson H V, Searle S R. The Vec-permutation Matrix, the Vec Operator and Kronecker Products: A Review. Linear and Multilinear Algebra, 1981

[28] Zha H. The Restricted Singular Value Decomposition of Matrix Triplets. SIAM J. Matrix Anal. Appl, 1991

[29] 李乔.矩阵论八讲.上海:上海科学技术出版社,1988

[30] 姜家辉.矩阵理论基础.大连:大连理工大学出版社,1997

[General Information]
书名=矩阵分析
作者=冯良贵，胡庆军编著
页数=154
SS号=12665886
DX号=
出版日期=2010.06
出版社=国防科技大学出版社